LOGIC, LANGUAGE
AND
COMPUTATION

CSLI
Lecture Notes
No. 58

LOGIC, LANGUAGE
AND
COMPUTATION
Volume 1

edited by
Jerry Seligman
&
Dag Westerståhl

CSLI Publications
CENTER FOR THE STUDY OF
LANGUAGE AND INFORMATION
STANFORD, CALIFORNIA

Printed in the United States
99 98 97 96   5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Logic, language and computation / edited by Jerry Seligman & Dag Westerståhl.
p. cm. — (CSLI lecture notes ; no. 58)
Includes bibliographical references.
ISBN 1-881526-90-9 (v. 1). — ISBN 1-881526-89-5 (pbk.).
1. Language and logic. 2. Logic, Symbolic and mathematical. 3 Computational linguistics. I. Seligman, Jerry, 1964– . II. Westerståhl, Dag, 1946– . III. Series.
P39.L593 1995
410'.285 — dc 2095-50832
CIP

CSLI was founded early in 1983 by researchers from Stanford University, SRI International, and Xerox PARC to further research and development of integrated theories of language, information, and computation CSLI headquarters and CSLI Publications are located on the campus of Stanford University

CSLI Lecture Notes report new developments in the study of language, information, and computation In addition to lecture notes, the series includes monographs, working papers, and conference proceedings Our aim is to make new results, ideas, and approaches available as quickly as possible

The acid-free paper used in this book meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Contents

1

Generalised Set Theory

PETER ACZEL

Introduction

One of the aims of work in STASS has been the development of a rigorous mathematical Situation Theory. This has been problematic because of the use, in situation theory, of parameters and a rich variety of parametric structured objects, including sets. The parameters in a parametric object can be substituted for and can be abstracted over or quantified over to form such objects as relations, types or propositions. A mathematical theory was needed that combined standard semantical notions like set with ideas, coming largely from syntax, to do with parameters. One approach to this challenge has been developed over a series of papers, Aczel 1990, Aczel and Lunnon 1991, Lunnon 1991a, Lunnon 1991b. This algebraic approach formulated a series of mathematical notions of universe of structured objects that incorporated more and more of the general features wanted by a universe of objects for situation theory. On the basis of these papers it seems that we now have the mathematical tools needed to cope with all the new kinds of object that seem to be needed in situation theory.

The aim of the present paper is to describe a set theoretical metatheory that incorporates the new kinds of objects in pure form. The hope is that the new metatheory will have an intuitive appeal and can be used informally in applications of mathematics as easily as the standard set theoretical metatheory, usually formalised as ZFC. In particular the hope is to use it in the description of a universe of objects for situation theory, without any need for the algebraic apparatus.

This paper is a major revision of a draft paper written in 1990, immediately after the second STASS meeting at Loch Rannoch. That draft paper was written out of a feeling of dissatisfaction with the ever growing complexity of the algebraic approach I had been taking.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

We do not yet want to commit ourselves to a specific body of axioms for the new metatheory, as we expect this to evolve with experience. We will call any metatheory developed along the lines presented in this paper a *Generalised Set Theory*. A starting point for us is the basic generalised set theory GST_0, a modification of ZFC to allow for both non-well-founded sets and non-sets.

Basic generalised set theory, GST_0

It is a well known fact that the universe of pure well founded sets, that gives the standard interpretation of the axiomatic set theory ZFC, is rich enough to represent all the mathematical objects of classical mathematics via suitable codings. It is this fact that is one of the reasons for the status of axiomatic set theory in modern mathematics as a standard foundational framework for mathematics.

But in spite of this fact it may be worthwhile to consider the development of mathematics in a larger universe including other objects besides the pure well founded sets. One direction to go is to drop the foundation axiom, FA, and perhaps replace it by an axiom asserting the existence of non-well-founded sets. This direction was explored in my book, Aczel 1988, where the anti-foundation axiom AFA or variants of it were added to the axiom system $ZFC^- = ZFC - FA$.

Another direction to go is to drop the need for all objects to be sets. This involves an extension of the language of set theory with a new unary predicate to distinguish sets from non-sets and a slight modification of the non-logical axioms of ZFC so as to express, for example, that only sets can have elements and that, in the extensionality axiom, it is only sets with the same elements that must be equal. Uses of this kind of set theory may already be found in Barwise 1975.

But the two directions may be combined into one where there may be both non-sets and non-well-founded sets. Let us call the resulting set theory GST_0. It will be the basis for a number of set theories obtained by possibly adding new non-logical symbols and new axioms for them. An interpretation of GST_0 has been exploited in Barwise and Etchemendy 1987 where a (hyper-)universe, V_A, is used which consists of possibly non-well-founded sets built up from atoms taken from the class A of atoms.[1] This universe models GST_0 and also satisfies the modification of the anti-foundation axiom that allows for atoms. But Barwise and Etchemendy do not stop with just this universe but use an extension universe $V_A[X]$ as an auxilliary tool. This universe is an extension of V_A which has sets which may involve ob-

[1]The atoms are not elements of V_A, so that the whole universe involved is really $A \cup V_A$. In this paper my universes will contain all the relevent objects, except in the following discussion about V_A and $V_A[X]$

jects called indeterminates from the class X as well as the atoms from the class A. Really the indeterminates in X are just new atoms, but have a different role to play in the applications. As well as there being a precise notion of an indeterminate *occuring* in a set there is a precise notion of *substitution* of sets from V_A (or atoms from A) for the indeterminates occuring in a set of $V_A[X]$ to give a set of V_A. (The definition of substitution makes essential use of the anti foundation axiom that holds in V_A.) Using substitution they are able to formulate a powerful tool for exploiting the anti-foundation axiom. This is the solution lemma. The lemma states that every system of equations over $V_A[X]$ has a unique solution in V_A. Here a system of equations over $V_A[X]$ has the form

$$x = a_x \quad (x \in X)$$

where each a_x is in $V_A[X]$. A solution in V_A to such a system of equations is an assignment s of a set $s_x \in V_A$ to each $x \in X$ such that

$$s_x = s * a_x \text{ for all } x \in X,$$

where, for any $a \in V_A[X]$, $s * a$ is the set of V_A obtained from a by substituting s_x for each occurence of x in a for each $x \in X$. In their book Barwise and Etchemendy study some situation theoretic objects that are modelled in V_A. The larger universe $V_A[X]$ is only used as a convenient tool via the solution lemma. But it is also of interest to focus on $V_A[X]$ itself. Following the situation theory tradition we will now call the elements of X *parameters* and the sets in $V_A[X]$ that involve at least one parameter *parametric* sets and the others *non-parametric* sets. Parameters and parametric objects of one kind or another are of explicit interest in situation theory and are being exploited in situation semantics. See, for example, Gawron and Peters 1990. We might hope to use $V_A[X]$ to model these situation theoretic parametric objects in the same kind of way that certain non-parametric objects of situation theory were modelled, using V_A, by Barwise and Etchemendy in their book.

Not only are parameters and parametric objects used in situation theory but also various kinds of abstractions of parameters are used, analogous to variable binding in the syntax of formal languages. In order to be able to easily model abstraction of parameters in our set theoretic models for situation theory we aim to build in parameter abstraction into our set theory. This means that among the non-sets of our set theory there will be objects called abstracts of the form $[\overline{x}]a$ where a is a possibly parametric object of the set theory and $\overline{x}$ is a non-repeating family of the parameters that may occur (free) in a but get bound in the abstract.

Outline of the paper

In section 1 we develop a general theory of parametric objects. Our first notion will be that of a *regular normal Substitution System*, which is intended to capture the idea of a universe of possibly parametric objects having a good notion of substitution into any object of the universe for the parameters that may occur in the object. The notion of a substitution system is very closely related to several other earlier notions, the earliest being Anil Nerode's notion of a *compositum*, Nerode 1956. See also Aczel 1991 and Barthe 1993, where further references may be found.

Next we will formulate the notions of an *abstraction operation* for such a universe and an *application operation* compatible with it to get the notion of a *Lambda Substitution System*. These give rise to models of the untyped lambda calculus which behave like open term models which use all possibly open terms, not just the closed terms that have no free variables. We also formulate, for any set I, I-indexed versions of the notions of abstraction and application operations. In an I-indexed abstraction an I-indexed non-repeating family of parameters are simultaneously abstracted over, and correspondingly, in an I-indexed application an object must be applied to an I-indexed family of objects.

In section 2 the general theory of the previous section is applied to describe informally a series of generalised set theories, ending with GST_4, in which the universe forms a Lambda Substitution System having I-indexed abstractions and application operations for all sets I. How do we know that these generalised set theories are relatively consistent to the standard axiomatic set theory ZFC? For GST_1 and GST_2 with suitable anti-foundation axioms the paper Aczel 1990 is a sufficient reference. For GST_3 we need the ideas developed in Aczel and Lunnon 1991, Lunnon 1991b and Lunnon 1991a. Finally, in order to model GST_4 the further ideas in Lunnon 1994 are needed.

In section 3 we describe an extension of GST_4 which uses ideas from Aczel 1980 to give an approach to an unsituated internal logic for situation theory. We end with a brief discussion of *restricted parameters.*

A Primitive Notion of Indexed Family

In order to deal with I-indexed notions we will have to work with I-indexed families $\overline{a} = (a_i)_{i \in I}$. Such indexed families could perhaps be represented set theoretically as functions with domain I in the usual way; i.e. as certain sets of ordered pairs, with ordered pairs defined in the standard set theoretical way. But this may cause problems when substituting into families, when the index set I is parametric. This is because intuitively we do not want to substitute into the index set or into the indices in the index set. This is not a serious problem as we need only restrict the use of indexed families to those where the index set is non-parametric. But rather than do that it has seemed to me natural to experiment with an alternative approach to indexed families in which we treat them as primitive in our generalised set theory, rather than have them represented set theoretically. That there is no problem in doing this has been demonstrated in the earlier papers which develop the theory of structured objects.

We work informally in the set theory GST_0. We shall describe informally an extension GST_1 which has a primitive notion of indexed family. In GST_1 there is given a class *Fam* of non-sets called *families*. With each family $\overline{a}$ is associated a set I, called the *index set* of $\overline{a}$, and an assignment of an object a_i to each $i \in I$, called the *component of* $\overline{a}$ *at index* i. The following property is required to hold:-

For any set I *and any assignmment of an object* a_i *to each* $i \in I$ *there is a uniquely determined family* $\overline{a}$ *having index set* I *and having component* a_i *at each index* $i \in I$. *We then write*

$$\overline{a} = (a_i)_{i \in I}.$$

Note the following extensionality criterion for equality of families $\overline{a} = (a_i)_{i \in I}$ and $\overline{b} = (b_j)_{j \in J}$:-

$$\overline{a} = \overline{b} \iff [I = J \text{ and } a_i = b_i \text{ for all } i \in I].$$

For any class A let A^I be the class of all I-indexed families $\overline{a} = (a_i)_{i \in I}$ whose components a_i are all in A. We will call a family $\overline{a} = (a_i)_{i \in I}$ *non-repeating* if, for all $i, j \in I$, with $i \neq j$

$$a_i \neq a_j.$$

The class of all non-repeating $\overline{a} \in A^I$ will be written $A^{\cdot I}$.

1.1 A Theory of Parametric Objects

1.1.1 Substitution Systems

Definition 1 *A* substitution system $\mathbb{A} = (A, X, *)$ *consists of a class* A, *a subclass* X *of* A *and an infix operation* $- * - : Sub \times A \to A$, *where* Sub *is the class of all partial functions from* X *to* A. *The conditions 1-4, given below, must hold.*

We will write [] *for the completely undefined element of* Sub. *If* $s_1, s_2 \in Sub$ *then we let* $s_1 * s_2 \in Sub$ *be the partial function with domain* $dom(s_1) \cup dom(s_2)$ *such that for all* $y \in dom(s_1 * s_2)$

$$(s_1 * s_2)y = s_1 * (s_2 * y).$$

If Y *is a subset of* X *and* $a \in A$ *then we write that* Y supports a *if for all* $s_1, s_2 \in Sub$

$$(\forall y \in Y)\ (s_1 * y = s_2 * y) \implies s_1 * a = s_2 * a.$$

We think of the elements of X as the parameters *of the system and the elements of A as the possibly parametric objects. We will call the elements of Sub the* substitutions *of the system, and for $s \in Sub, a \in A$ we think of $s * a$ as the result of 'applying' the substitution to a; i.e. the result of simultaneously substituting in a the object sx for each parameter x in the domain of s.*

We now list the four conditions for $\mathbb{A}$ to be a substitution system:-

1. *For all $s \in Sub, x \in X$*
$$s * x = \begin{cases} sx & \text{if } x \in dom(s) \\ x & \text{otherwise.} \end{cases}$$
2. *$[] * a = a$ for all $a \in A$.*
3. *$s * (s' * a) = (s * s') * a$ for all $s_1, s_2 \in Sub, a \in A$.*
4. *Every element of A has a support set.*

If the following strengthening of 4 holds then we write that the substitution system is normal*:-*

4′. *Every element of A has a smallest support set.*

When this holds then, for each $a \in A$, we define $par(a)$ to be the uniquely determined smallest support set for a. We can write that a depends on *the parameters in $par(a)$, or alternatively, that the parameters in $par(a)$* occur in *a.*

Note that condition 4 above always holds when X is a set. But in a generalised set theory, where we will want the universe, a proper class, to form a normal substitution system, if there are also indexed abstractions then we will want an unlimited supply of parameters and this means that X has to be a proper class. Nevertheless we only want any object of the universe to depend on a *set* of parameters. So condition 4 is needed.

A substitution system is *finitary* if every element has a finite support. It is not hard to show that every finitary substitution system is normal using the following result.

Lemma 1 *For each element of a substitution system the intersection of any two support sets is also a support set.*

In general when working with (indexed) abstraction on a substitution system it is natural to require that 'there are enough parameters' for the same reason that in finitary formal languages such as the predicate calculus or in the lambda calculus, where variables can get bound, it is natural to assume that there are infinitely many variables, although only finitely many will occur in any given expression. The following definition captures this idea.

Definition 2 *A substitution system is* regular *if there is an infinite regular* *cardinal κ, that is less than or equal to the cardinality of X, such that every element of A has a support set of cardinality less than κ.*

We can include the case when X is a proper class in this definition by allowing $\kappa = \infty$. The smallest possible cardinal κ in the definition is called the *rank* of $\mathbb{A}$.

Notation: It will be convenient to define the substitution $[\overline{a}/\overline{x}]$, where $\overline{x}$ is in the class $X^{\ I}$ of all I-indexed non-repeating families of elements of X and $\overline{a}$ is in the class A^I of all I-indexed families of elements of A. We define it to be the substitution, defined on the set $C\overline{x} = \{x_i \mid i \in I\}$ of all components of $\overline{x}$, that maps x_i to a_i for each $i \in I$. We will usually write $[\overline{a}/\overline{x}]b$ rather than $[\overline{a}/\overline{x}] * b$. The special case when $I = \{1, \dots, n\}$ can be written $[a_1, \dots, a_n/x_1, \dots, x_n]b$.

1.1.2 Solving equations and systems of equations

An *equation* over a substitution system $\mathbb{A} = (A, X, *)$ is given by a pair $(x, a) \in X \times A$ and will usually be written

$$x = a.$$

Note that a may depend on x so that the equation is recursive. A *solution* to the equation is an element $b \in A$ such that

$$b = [b/x]a.$$

If every equation has a solution then I write that the substitution system has the *Equation Solving Property.*

We now formulate the indexed version of these notions. An *I-indexed system of equations*, abbreviated *I-soe*, over the substitution system is given by a pair $(\overline{x}, \overline{a}) \in X^{\ I} \times A^I$ and is written

$$\overline{x} = \overline{a}$$

or sometimes

$$x_i = a_i \quad (i \in I).$$

A *solution* to the I-soe is a family $\overline{b}$ such that, for all $i \in I$,

$$b_i = [\overline{b}/\overline{x}]a_i.$$

This can be written more concisely as

$$\overline{b} = [\overline{b}/\overline{x}]\overline{a}$$

if we take the right hand side to abbreviate $([\overline{b}/\overline{x}]a_i)_{i \in I}$. If every I-soe has a solution then I write that the substitution system has the *I-soe Solving Property.*

Note the similarities and differences from the Solution Lemma, which can be put in the general form:-

Every guarded system of equations has a unique solution.

Some notion of guardedness is essential to ensure a unique solution. For example any element will be a solution of the equation

$$x = x,$$

so that this cannot be allowed to be a guarded equation. The solution lemma has two aspects. First the existence of a solution and second its uniqueness. The Solving Property gives us the existence part of the Solution Lemma, but not uniqueness.

1.1.3 Ordinary Abstraction

We assume given a regular normal substitution system $\mathbb{A}$. We wish to capture formally the following natural informal notion in connection with parametric objects. Given an object a and a parameter x which may occur in a we can imagine removing all occurrences of x in a if any from a leaving *holes* in the places in a where previously x occurred. Let us write $[x]a$ for the resulting entity and call it an *abstract*. Note that the parameters other than x that occur in a should still occur in the abstract and it should still be possible to substitute for those parameters. The abstract $[x]a$ was obtained from the pair (x, a). The same abstract may sometimes also be obtained from another pair (y, b). When will this happen; i.e. when is

$$[x]a = [y]b?$$

We now give three natural candidate conditions for this.

1. $b = [y/x] * a$ and $y \notin par(a) - \{x\}$.
2. $[z/x] * a = [z/y] * b$ for some $z \in X - (par(a) \cup par(b))$.
3. $[c/x] * a = [c/y] * b$ for all $c \in A$.

Fortunately we have:-

Proposition 2 *Conditions 1,2,3 above are equivalent.*

If any of these conditions hold we write that

$$(x, a) \sim (y, b).$$

It is clear that $\sim$ is an equivalence relation on $X \times A$. The following result is useful:-

Proposition 3 *If* $(x, a) \sim (y, b)$ *then*

1. $par(a) - \{x\} = par(b) - \{y\}$,
2. $(x, s*a) \sim (y, s*b)$ *for any substitution* s *such that* $s*x = x, s*y = y$ *and* $x, y \notin par(s * z)$ *for all* $z \in par(a) - \{x\}$.

As in the lambda calculus we want to consider the possibility that abstractions of objects of the substitution system $\mathbb{A}$ are themselves such objects. But then we have to consider how the substitutions should interact with the abstractions. By examining the properties of substitution in the lambda calculus we are led to the following definition:-

Definition 3 *An operation* $\lambda : X \times A \to A$ *is an* abstraction operation *on* $\mathbb{A}$ *if*

1. *For all* $(x, a), (y, b) \in X \times A$
$$\lambda(x, a) = \lambda(y, b) \iff (x, a) \sim (y, b),$$
2. *For all* $(x, a) \in X \times A$
$$par(\lambda(x, a)) = par(a) - \{x\},$$
3. *For all* $(x, a) \in X \times A$ *and* $s \in Sub$ *if* $x \notin par(s * y)$ *for all* $y \in Y = par(a) - \{x\}$ *then*
$$s * \lambda(x, a) = \lambda(x, (s \restriction Y) * a).$$

Note: I expect that 2 is redundant.

In the lambda calculus there is also an application operation satisfying the standard beta equality law. We are led to the following notion:-

Definition 4 *Given an abstraction operation* λ *on* $\mathbb{A}$, *an* application operation for λ *is an infix operation* $-@- : A \times A \to A$ *such that*

1. *For all* $(x, a) \in X \times A, b \in A$
$$\lambda(x, a)@b = [b/x]a,$$
2. *For all* $a, b \in A, s \in Sub$
$$s * (a@b) = (s * a)@(s * b).$$

We shall call an abstraction system, together with an application operation for it, a *lambda system* for $\mathbb{A}$.

Notation: We shall continue to write $[x]a$ for $\lambda(x, a)$, when convenient, when working with an abstraction operation. Also, when working with an application operation @ we shall follow the standard practice of keeping the operation implicit; i.e. writing just ab for $a@b$. Also, as usual, $ab_1 \cdots b_n$ will abbreviate $(\cdots(ab_1)\cdots)b_n$.

Note: When $\mathbb{A}$ is finitary then a lambda system for it gives rise to a 'model of the lambda calculus' in a very simple way. There are various notions of 'model of the lambda calculus' in the literature. I believe that the notion of a regular finitary substitution system with a lambda system corresponds to the notion of a λ-algebra. But we leave the exploration of this for another publication.

1.1.4 Indexed Abstraction

We now want to give an indexed version of the previous subsection. As there, we assume given a regular normal substitution system $\mathbb{A}$. In the finite case we may wish to form abstracts $[x_1, \ldots, x_n]a$ having $n > 1$ kinds of possible holes after removing possible occurences of n distinct parameters $x_1, \ldots, x_n$ in a. More generally we may consider abstracts $[\overline{x}]a$ where, for some set I, $\overline{x} = (x_i)_{i \in I}$ is in the class $X^{\cdot I}$ of non-repeating I-indexed families of parameters. The abstract $[x_1, \ldots, x_n]a$ is now the special case when $I = \{1, \ldots, n\}$. In this subsection we assume that I is a fixed set of cardinality less than the rank of $\mathbb{A}$.

As with ordinary abstraction we give the candidate conditions for pairs $(\overline{x}, a), (\overline{y}, b) \in X^{\cdot I} \times A$ to give rise to abstracts $[\overline{x}]a, [\overline{y}]b$ that are equal.

1. $b = [\overline{y}/\overline{x}] * a$ and $y_i \notin par(a) - C\overline{x}$ for all $i \in I$.
2. $[\overline{z}/\overline{x}] * a = [\overline{z}/\overline{y}] * b$ for some $\overline{z} \in Y^{\cdot I}$, where $Y = (X - (par(a) \cup par(b))$.
3. $[\overline{c}/\overline{x}] * a = [\overline{c}/\overline{y}] * b$ for all $\overline{c} \in A^I$.

Again we have the result:

Proposition 4 *The conditions 1,2,3 are equivalent.*

If any of the equivalent conditions 1,2,3 hold then we write

$$(\overline{x}, a) \sim_I (\overline{y}, b).$$

Using 3 it is clear that $\sim_I$ is an equivalence relation on $X^{\cdot I} \times A$. The following result will be useful.

Proposition 5 *If $(\overline{x}, a) \sim_I (\overline{y}, b)$ then*

1. *$par(a) - C\overline{x} = par(b) - C\overline{y}$,*
2. *$(\overline{x}, s * a) \sim_I (\overline{y}, s * b)$ for any substitution s such that $s * u = u$ for all components u of $\overline{x}$ and $\overline{y}$ and non of those components is in $par(s * z)$ for any $z \in par(a) - C\overline{x}$.*

We are now ready to define indexed abstractions and applications.

Definition 5 *An operation $\lambda^I : X^{\cdot I} \times A \to A$ is an I-*indexed abstraction operation *on $\mathbb{A}$ if*

1. *For all $(\overline{x}, a), (\overline{y}, b) \in X^{\cdot I} \times A$*
$$\lambda^I(\overline{x}, a) = \lambda^I(\overline{y}, b) \iff (\overline{x}, a) \sim_I (\overline{y}, b),$$
2. *For all $(\overline{x}, a) \in X^{\cdot I} \times A$*
$$par(\lambda^I(\overline{x}, a)) = par(a) - C\overline{x},$$
3. *For all $(\overline{x}, a) \in X^{\cdot I} \times A$ and $s \in Sub$ if no component of $\overline{x}$ is in $par(s * y)$ for any $y \in Y = par(a) - C\overline{x}$ then*
$$s * \lambda^I(\overline{x}, a) = \lambda^I(\overline{x}, (s \restriction Y) * a).$$

Note: I expect that 2 is redundant.

Definition 6 *Given an I-indexed abstraction operation λ^I on $\mathbb{A}$, an* application operation *for it is an infix operation $-@^I- : A \times A^I \to A$ such that*

1. *For all $(\overline{x}, a) \in X^{\ I} \times A, \overline{b} \in A^I$*

   $$\lambda^I(\overline{x}, a)@\overline{b} = [\overline{b}/\overline{x}]a,$$

2. *For all $a \in A, \overline{b} \in A^I, s \in Sub$*

   $$s * (a@^I\overline{b}) = (s * a)@^I(s * b_i)_{i \in I}.$$

We shall call an I-indexed abstraction system, together with an application operation for it, an *I-indexed lambda system* for $\mathbb{A}$.

Notation: We shall continue to write $[\overline{x}]a$ for $\lambda^I(\overline{x}, a)$, when convenient, when working with an abstraction operation. Also, when working with an application operation $@^I$ we shall write just $a\overline{b}$ for $a@^I\overline{b}$.

Some further abbreviations will also be useful. Given sets I, J, if $\overline{b} \in A^J, \overline{x} \in X^{\ J}, \overline{a} = (a_i)_{i \in I} \in A^I$, let

$$\begin{cases} [\overline{b}/\overline{x}]\overline{a} & = ([\overline{b}/\overline{x}]a_i)_{i \in I} \\ \overline{a}\overline{b} & = (a_i\overline{b})_{i \in I} \\ [\overline{x}]\overline{a} & = ([\overline{x}]a_i)_{i \in I} \end{cases}$$

Note that, using these abbreviations, we have the following indexed version of beta equality:-

$$([\overline{x}]\overline{a})\overline{b} = [\overline{b}/\overline{x}]\overline{a}$$

1.1.5 Solving Equations using Lambda Systems

We show that if a regular normal substitution system has a lambda system then it has the equation solving property. The argument carries over to the indexed version

Theorem 6 *Let $\mathbb{A} = (A, X, *)$ be a regular normal substitution system.*

1. *If $\mathbb{A}$ has a lambda system on it then it has the equation solving property.*
2. *If $\mathbb{A}$ has an I-indexed lambda system on it, where I is a set of cardinality less than the rank of $\mathbb{A}$, then it has the I-soe solving property.*

Proof:

1. Given $(x, a) \in X \times A$ let $c = ee$ where $e = [x][xx/x]a$. Then

   $$c = ee = [ee/x]a = [c/x]a.$$

2. Given $(\overline{x}, \overline{a}) \in X^{\ I} \times A^I$ let $\overline{c} = \overline{e}\overline{e}$ where $\overline{e} = [\overline{x}][\overline{x}\overline{x}/\overline{x}]\overline{a}$. Then

   $$\overline{c} = \overline{e}\overline{e} = [\overline{e}\overline{e}/\overline{x}]\overline{a} = [\overline{c}/\overline{x}]\overline{a}.$$

Note that in 1 we could have let $c = Ya$, using the standard fixed-point combinator

$$Y = [y](\ ([x](y(xx)))\ ([x](y(xx)))\ ).$$

The indexed version $\overline{Y}$ can be defined and used to get $\overline{c}$ in 2.

1.2 The Generalised Set Theories GST_i for $i = 2, 3, 4$

Recall that GST_0 is the axiomatic set theory ZFC^-, modified so as to allow for non-sets. Also GST_1 is obtained from GST_0 by postulating a primitive notion of indexed family. In particular, in GST_1 there is a class Fam of non-sets called families and assignments of an index set and components to each family satisfying a suitable property. We now want to successively add the notions of our theory of parametric objects to GST_1; i.e. a regular normal substitution system, ordinary and indexed abstractions on the substitution system and finally application operations for them.

We use V for the universal class of all objects and Set for the class of all sets.

GST_2

This theory is obtained from GST_1 by postulating classes Atm and Par of *atoms* and *parameters* and an operation $- * - : Sub \times V \to V$, where Sub is the class of all functions with domain a set of parameters. These notions must satisfy the following conditions:-

1. Atm and Par are disjoint from each other and from Set and Fam.
2. Par is a proper class and $\mathbb{V} = (V, Par, *)$ forms a regular normal substitution system.
3. For any substitution $s \in Sub$
   a. $s * a = a$ for all $a \in Atm$.
   b. $s * a = \{s * b \mid b \in a\}$ for all $a \in Set$.
   c. $s * \overline{a} = (s * a_\imath)_{\imath \in I}$ for all $\overline{a} = (a_\imath)_{\imath \in I} \in Fam$.

GST_3

This theory is obtained from GST_2 by postulating an operation $\lambda : Par \times V \to V$ and, for each set I, an operation $\lambda^I : Par^{\,I} \times V \to V$. These must satisfy the following conditions, where $B = \{\lambda(x, a) \mid (x, a) \in Par \times V\}$ and, for each set I, $B_I = \{\lambda^I(\overline{x}, a) \mid (\overline{x}, a) \in Par^{\,I} \times V\}$:-

1. λ is an abstraction operation for $\mathbb{V}$ and , for each set I, λ^I is an I-indexed abstraction operation for $\mathbb{V}$.
2. The class B and the classes B_I, for sets I, are all disjoint from Atm, Par, Set and Fam. Also B is disjoint from each B_I and the classes B_I are pairwise disjoint.

GST_4

This theory is obtained from GST_3 by postulating an operation $-@- : V \times V \to V$ and a distinguished atom $\bot \in Atm$ satisfying the following conditions:-

1. $-@-$ is an application operation for λ.
2. For each set I $-@^I- : V \times V^I \to V$ is an application operation for λ^I where, for $a \in V, \bar{b} \in V^I$,'

   $$a@^I\bar{b} = a@\bar{b}.$$

3. For all $b \in V$

   $$\bot@b = \bot \text{ and } a@b = \bot \text{ for all } a \in Atm \cup Set \cup Fam.$$

4. For all $a, b \in V$ and all bisimulations R

   $$aRb \implies a = b.$$

In the last condition the notion of bisimulation relation is defined as follows. A relation R is a *bisimulation relation* if

$$R \subseteq R^{+Atm} \cup R^{+Set} \cup R^{+Fam} \cup R^{+\lambda} \cup R^{+@},$$

where

$$aR^{+Atm}b \iff a, b \in Atm \;\&\; a = b,$$

$$aR^{+Set}b \iff a, b \in Set \;\&\; \forall x \in a \exists y \in b\; xRy \;\&\; \forall y \in b \exists x \in a\; xRy,$$

$$aR^{+Fam}b \iff a, b \in Fam \;\&\; \text{ for some set } I \\ a = (a_\imath)_{\imath \in I}, b = (b_\imath)_{\imath \in I} \;\&\; \forall \imath \in I\; a_\imath R b_\imath,$$

$$aR^{+\lambda}b \iff \exists x \in Par\;\; \exists a', b' \\ [a = \lambda(x, a') \;\&\; b = \lambda(x, b') \;\&\; a'Rb'] \\ \text{or for some set } I \;\exists \bar{x} \in Par^{\,I}\;\; \exists a', b' \\ [a = \lambda^I(\bar{x}, a') \;\&\; b = \lambda^I(\bar{x}, b') \;\&\; a'Rb'],$$

$$aR^{+@}b \iff \exists x \in Par\;\; \exists a_1, \ldots, a_n \exists b_1, \ldots, b_n\; [a_\imath R b_\imath \text{ for } i = 1, \ldots, n \\ \;\&\; a = ((\cdots(x@a_1)@\cdots)@a_n) \;\&\; b = ((\cdots(x@b_1)@\cdots)@b_n)].$$

Condition 4 is a strong extensionality axiom which, when combined with the I-soe property for all sets I, we expect can be used to prove an appropriate anti-foundation property and even a solution property of the standard form:-

For any set I, every guarded I-soe has a unique solution.

We do not have space to consider more details here. See Lunnon 1994 for an anti-foundation axiom for Lunnon's generalised set theory $ZF\lambda$.

1.3 Unsituated Logic in Generalised Set Theory

We believe that a mathematical situation theory is best understood by developing its conceptual ingredients in three layers:

1. Pure Ontology
2. Unsituated Logic
3. Situated Logic

The generalised set theory GST_4 is intended to give us the first layer. It supplies the basic forms of objects that seem to be needed. We conclude the paper with a brief summary of an approach to the second layer. At this layer we want to deal with the notions of *proposition, true proposition, type* and *restricted parameters.* There are undoubtedly a variety of approaches that can be taken, and have been taken, to the development of an unsituated logic for situation theory. Here we take a stand that rules out some of the earlier approaches:-

> *Only non-parametric objects should be first class in the internal logic of situation theory.*

So, for example, we are ruling out having a type of parameters as a type of the internal logic. Of course we do have the class *Par* of all parameters and parameters and parametric objects are certainly first class in the generalised set theory GST_4. But we should not expect the statements of generalised set theory to be generally expressible in the internal logic of situation theory. The point is that we want to use a generalised set theory only as a *metatheory* for the internal logic of situation theory.

We extend the theory GST_4 by postulating classes *Prop* and *True* and distinguished objects, called *logical constants*, $\doteq, \dot{\bigwedge}, \dot{\supset}, \dot{\neg}, \dot{\pi}, \dot{\forall}$ satisfying the conditions:-

1. $True \subseteq Prop \subseteq Obj$, where Obj is the class of all non-parametric objects.
2. The logical constants are distinct atoms satisfying:-

The Logical Schemata

Equality If $a, b \in Obj$ then $(\doteq ab) \in Prop$ and
$$(\doteq ab) \in True \iff a = b.$$

Conjunction If Φ is a subset of $Prop$ then $(\dot{\bigwedge}\Phi) \in Prop$ and
$$(\dot{\bigwedge}\Phi) \in True \iff \Phi \subseteq True.$$

Negation If $\phi \in Prop$ then $(\dot{\neg}\phi) \in Prop$ and
$$(\dot{\neg}\phi) \in True \iff \phi \notin True.$$

Strong Implication If $\phi \in Prop$ and $\psi \in Obj$ such that either $\phi \notin True$ or $\psi \in Prop$ then $(\dot{\supset}\phi\psi) \in Prop$ and

$$(\dot{\supset}\phi\psi) \in True \iff [\phi \notin True \text{ or } \psi \in True].$$

Predication If $t \in Type$ and $a \in Obj$ then $(\dot{\pi}ta) \in Prop$ and

$$(\dot{\pi}ta) \in True \iff (ta) \in True.$$

Universal Quantification If $t \in Type$ then $(\dot{\forall}t) \in Prop$ and

- if t is an ordinary type then
$$(\dot{\forall}t) \in True \iff (tb) \in True \text{ for all } b \in Obj,$$
- if, for some set I, t is an I-indexed type then
$$(\dot{\forall}t) \in True \iff (t\bar{b}) \in True \text{ for all } \bar{b} \in Obj^I.$$

In the last two schemata we have used the notion of a type. We need the following definitions. An *ordinary type* is a non-parametric ordinary abstraction $[x]a$, where $(x, a) \in Par \times V$ such that $[b/x]a \in Prop$ for all $b \in Obj$. For a given set I, an *I-indexed type* is an I-indexed abstraction $[\bar{x}]a$, where $(\bar{x}, a) \in Par^I \times V$ such that $[\bar{b}/\bar{x}]a \in Prop$ for all $\bar{b} \in Obj^I$. We write *Type* for the class of all types, ordinary or I-indexed for some set. This is a notion of *total type*. There is also an apparent need for a notion of *partial type* in situation theory, where partial types may have non-vacuous *appropriateness conditions*. We leave a discussion of this and the development of a situated logic in generalised set theory for a future paper.

Restricted Parameters

Restricted parameters, and objects parametric in them, seem to be very useful in situation semantics. See for example Gawron and Peters 1990. But what are they and how should we treat them in our unsituated logic?

Roughly a restricted parameter x^ϕ is a parameter x, governed by a constraint ϕ. It seems to be a special kind of entity that we have not allowed for so far in our generalised set theory. In my view it is unneccessary to have these entities and statements involving them should be considered to be possibly convenient rephrasings of statements not involving them.

We only have space for a simple example to illustrate the idea. Let $x \in Par$ and $\phi, \psi \in V$ such that $par(\phi) = par(\psi) = \{x\}$. Suppose that ϕ is a *parametric proposition*. By this we mean that $[x]\phi$ is an ordinary type; i.e. $[a/x]\phi \in Prop$ for all $a \in Obj$. Then we can use ϕ as a constraint in forming the restricted parameter x^ϕ and substitute it for x in ψ to obtain ψ', a parametric object depending on the restricted parameter x^ϕ. Now we may wish to assert the statement that

(*) $\quad\psi'$ is a parametric proposition

What should this mean? Surely it must mean that

$$(**) \quad [a/x]\psi \in Prop \text{ for all } a \in Obj \text{ such that } [a/x]\phi \in True.$$

Note that (**) does not involve restricted parameters. So we can take (*) to mean nothing else than a rephrasing of (**). I expect that all statements involving restricted parameters can be treated in this way. Also note that if (**) holds then so does

$$(***) \quad (\dot{\supset}\phi\psi) \text{ is a parametric proposition.}$$

Perhaps we should take (*) to be a rephrasing of (***).

References

Aczel, Peter. 1980. Frege structures and the notions of proposition, truth and set. In *The Kleene Symposium*, ed. H. Jerome Keisler Jon Barwise and Keneth Kunnen. Studies in Logic and the Foundations of Mathematics, No. 101. Amsterdam, New York, Oxford: North-Holland.

Aczel, Peter. 1988. *Non-well-founded Sets.* CSLI Lecture Notes Number 14. Stanford: CSLI Publications.

Aczel, Peter. 1990. Replacement Systems and the Axiomatization of Situation Theory. In *Situation Theory and Its Applications, I*, ed. Robin Cooper, Kuniaki Mukai, and John Perry. CSLI Lecture Notes, No. 22. Stanford: CSLI Publications.

Aczel, Peter. 1991. Term Declaration Logic and Generalised Composita. In *Proceedings of the Sixth IEEE Symposium on Logic in Computer Science.*

Aczel, Peter, and Rachel Lunnon. 1991. Universes and Parameters. In *Situation Theory and Its Applications, II*, ed. Jon Barwise, J. M. Gawron, G. Plotkin, and S. Tutiya. CSLI Lecture Notes, No. 26. Stanford: CSLI Publications.

Barthe, Gilles. 1993. *Term Declaration Logic and Generalised Composita.* Doctoral dissertation, Manchester University, Manchester, UK.

Barwise, Jon. 1975. *Admissible Sets and Structures.* New York: Springer-Verlag.

Barwise, Jon, and John Etchemendy. 1987. *The Liar: An Essay on Truth and Circularity.* Oxford: Oxford University Press.

Gawron, Jean Mark, and Stanley Peters. 1990. *Anaphora and Quantification in Situation Semantics.* CSLI Lecture Notes, No. 19. Stanford: CSLI Publications.

Lunnon, Rachel. 1991a. *Generalised Universes.* Doctoral dissertation, Manchester University, Manchester, UK.

Lunnon, Rachel. 1991b. Many Sorted Universes, SRDs and Injective sums. In *Situation Theory and Its Applications, II*, ed. Jon Barwise, J. M. Gawron, G. Plotkin, and S. Tutiya. CSLI Lecture Notes, No. 26. Stanford: CSLI Publications.

Lunnon, Rachel. 1994. An Anti-Founded Model of a Theory of Sets and Functions. Preprint.

Nerode, Anil. 1956. *Composita, Equations and Recursive Definitions.* Doctoral dissertation, The University of Chicago, Chicago.

2

Information-Oriented Computation with BABY-SIT

ERKAN TIN AND VAROL AKMAN

Introduction

While situation theory and situation semantics (Barwise and Perry 1983) provide an appropriate framework for a realistic model-theoretic treatment of natural language, serious thinking on their 'computational' aspects has only recently started (Black 1993, Nakashima et al. 1988). Existing proposals mainly offer a Prolog- or Lisp-like programming environment with varying degrees of divergence from the ontology of situation theory. In this paper, we introduce a computational medium (called BABY-SIT) based on situations (Tin and Akman 1994a, Tin and Akman 1994b). The primary motivation underlying BABY-SIT is to facilitate the development and testing of programs in domains ranging from linguistics to artificial intelligence in a unified framework built upon situation-theoretic constructs.

2.1 Constructs for Situated Processing

Intelligent agents generally make their way in the world as follows: pick up certain information from a situation, process it, and react accordingly (Devlin 1991, Dretske 1981, Israel and Perry 1990). Being in a (mental) situation, such an agent has information about the situations it sees, believes in, hears about, etc. Awareness of some type of situation causes the agent to acquire more information about that situation as well as other situation types, and to act accordingly. Assuming the possession of prior information and knowledge of some constraints, the acquisition of an item of information by an agent can also provide the agent with an additional item of information.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

Reaping information from a situation is not the only way an agent processes information. It can also act in accordance of the obtained information to change the environment. Creating new situations to arrive at new information and conveying information it already has to other agents are the primary functions of its activities.

In situation theory, abstraction can be captured in a primitive level by allowing parameters in infons. *Parameter-free infons* are the basic items of information about the world (i.e., 'facts') while parametric infons are the essential units that are utilized in a computational treatment of information flow.

To construct a computational model of situation theory, it is convenient to have available abstract analogs of objects. As noted above, by using parameters we can have parametric situations, parametric individuals, etc. This yields a rich set of data types. Abstract situations can be viewed as models of real situations. They are set-theoretic entities that capture only some of the features of real situations, but are amenable to computation. We define abstract situations as structures consisting of a set of parametric infons.

Information can be partitioned into situations by defining a hierarchy between situations. A situation can be larger, having other situations as its subparts. For example, an utterance situation for a sentence consists of the utterance situations for each word forming the sentence. The *part-of* relation of situation theory can be used to build hierarchies among situations and the notion of nested information can be accommodated.

Being in a situation, one can derive information about other situations connected to it in some way. For example, from an utterance situation it is possible to obtain information about the situation it describes. Accessing information both via a hierarchy of situations and explicit relationships among them requires a computational mechanism. This mechanism will put information about situation types related in some way into the comfortable reach of the agent and can be made possible by a proper implementation of the *supports* relation, $\models$, of situation theory.

Constraints enable one situation to provide information about another and serve as links. When viewed as a backward-chaining rule, a constraint can provide a channel for information flow between types of situations, from the antecedent to the consequent. This means that such a constraint behaves as a 'definition' for its consequent part. Another way of viewing a constraint is as a forward-chaining rule. This enables an agent to alter its environment.

2.2 Computational Situation Theory

2.2.1 PROSIT

PROSIT (PROgramming in SItuation Theory) is a situation-theoretic programming language (Schütze 1991, Nakashima et al. 1988). PROSIT is tailored more for general knowledge representation than for natural language processing. One can define situation structures and assert knowledge in particular situations. It is also possible to define relations between situations in the form of constraints. PROSIT's computational power is due to an ability to draw inferences via rules of inference which are actually constraints of some type. PROSIT can deal with self-referential expressions (Barwise and Etchemendy 1987).

One can assert facts that a situation should support and queries can be posed about one situation from another, but the results will depend on where the query is made.

Constraints can be specified as forward-chaining constraints, backward-chaining constraints, or both. Backward-chaining constraints are activated at query-time while forward-chaining constraints are activated at assertion-time. For a constraint to be applicable to a situation, the situation must be declared to 'respect' the constraint. Constraints in PROSIT are about local facts within a situation rather than about situation types. That is, the interpretation of constraints does not allow direct specification of constraints between situations, only between infons within situations.

Situated constraints offer an elegant solution to the treatment of *conditional constraints* which only apply in situations that obey some condition. This is actually achieved in PROSIT since information is specified in the constraint itself. Situating a constraint means that it may only apply to appropriate situations and is a good strategy to enforce *background conditions.* However, it might be required that conditions are set not only within the same situation, but also between various types of situations.

Parameters, variables, and constants are used for representing entities in PROSIT. Variables match any expression in the language and parameters can be equated to any constant or parameter. That is, the concept of *appropriateness conditions* is not exploited in PROSIT. It is more useful to have parameters that range over various classes rather than to work with parameters ranging over all objects.

Given a parameter of some type (individual, situation, etc.), an anchor is a function which assigns an object of the same type to the parameter (Devlin 1991, pp. 52–63). Hence, parameters work by placing restrictions on anchors. However, there is no appropriate anchoring mechanism in PROSIT since parameters are not typed.

2.2.2 ASTL

Black's ASTL (A Situation Theoretic Language) is another programming language based on situation theory (Black 1993). ASTL is aimed at natural language processing. The primary motivation underlying ASTL is to figure out a framework in which semantic theories can be described and possibly compared. One can define in ASTL constraints and rules of inference over the situations.

ASTL ontology incorporates individuals, relations, situations, parameters, and variables. These form the basic terms of the language. Situations can contain facts which have those situations as arguments. Sentences in ASTL are constructed from terms in the language and can be constraints, grammar rules, or word entries. Constraints are actually backward-chaining constraints and are global. Thus, a new situation of the appropriate type need not have a constraint explicitly added to it. Grammar rules are yet another sort of constraints with similar semantics. Although one can define constraints between situations in ASTL, the notion of a background condition for constraints is not available. Similar to PROSIT, ASTL cares little about coherence within situations. This is left to the user's control.

Declaring situations to be of some type allows abstraction over situations to some degree. But, the actual means of abstraction over objects in situation theory, viz., parameters, carry little significance in ASTL.

As in PROSIT, variables in ASTL have scope only within the constraint they appear. They match any expression in the language unless they are declared to be of some specific situation type in the constraint. Hence, it is not possible to declare variables (nor parameters) to be of other types such as individuals, relations, etc. Moreover, ASTL does not permit a definition of appropriateness conditions for arguments of relations.

ASTL does not have a mechanism to relate two situations so that one will support all the facts that the other does. This might be achieved via constraints, but there is no built-in structure between situations.

2.2.3 Situation Schemata

Situation schemata have been introduced (Fenstad et al. 1987) as a theoretical tool for extracting and displaying information relevant for semantic interpretation from linguistic form. A situation schema is an attribute-value system which has a choice of primary attributes matching the primitives of situation semantics. The boundaries of situation schemata are flexible and, depending on the underlying theory of grammar, are susceptible to amendment.

Situation schemata can be adopted to various kinds of semantic interpretation. One could give some kind of operational interpretation in a suitable programming language, exploiting logical insights. But in its present state, situation schemata do not go further than being a complex attribute-value structure. They allow representation of situations within this structure, but do not use situation theory itself as a basis. Situations, locations, individuals, and relations constitute the basic domains of the structure. Constraints are declarative descriptions of the relationships holding between aspects of linguistic form and the semantic representation itself.

2.3 BABY-SIT

2.3.1 Computational Model and Architecture

The computational model underlying the current version of BABY-SIT consists of nine primitive domains: *individuals* (I), *times* (T), *places* (L), *relations* (R), *polarities* (O), *parameters* (P), *infons* (F), *situations* (S), and *types* (K). Each primitive domain carries its own internal structure:

- Individuals: Unique atomic entities in the model which correspond to real objects in the world.
- Times: Individuals of distinguished type, representing temporal locations.
- Places: Similar to times, places are individuals which represent spatial locations.
- Relations: Various relations hold or fail to hold between objects. A relation has argument roles which must be occupied by appropriate objects.
- Polarities: The 'truth values' 0 and 1.
- Infons: Discrete items of information of the form $\ll rel,\ arg_1,\ \ldots,\ arg_n,\ pol \gg$, where *rel* is a relation, arg_i, $1 \leq i \leq n$, is an object of the appropriate type for the ith argument role, and *pol* is the polarity.
- Parameters: 'Place holders' for objects in the model. They are used to refer to arbitrary objects of a given type.
- Situations: (Abstract) situations are set-theoretic constructs, e.g., a set of *parametric infons* (comprising relations, parameters, and polarities). A parametric infon is the basic computational unit. By defining a hierarchy between them, situations can be embedded via the special relation *part-of.* A situation can be either (spatially and/or temporally) *located* or *unlocated.* Time and place for a situation can be declared by *time-of* and *place-of* relations, respectively.
- Types: Higher-order uniformities for individuating or discriminating uniformities in the world.

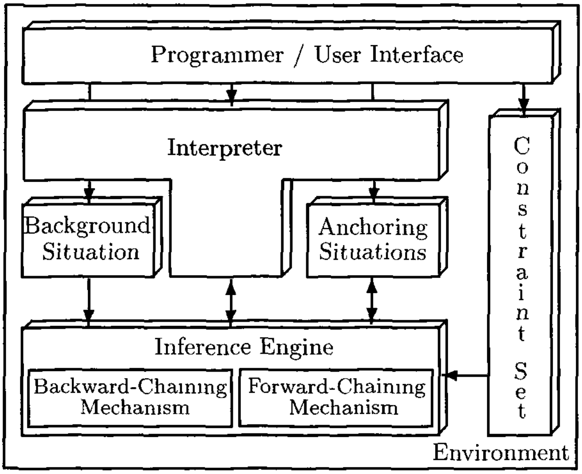


FIGURE 1 The Architecture of BABY-SIT.

The *model*, M, is a tuple $<I, T, L, R, O, P, F, S, K>$. This is shared by all components of the system. *Description* of a model, D_M, consists of a definition of M and a set of *constraints*, C. The *computational model* is then defined as a tuple $<D_M, A, A', U>$ where A is an anchor for parameters, A' is an assignment for variables, and U is an interpretation for D_M. A is provided by the anchoring situations while A' is obtained through unification. U is dynamically defined by the operational semantics of the computation. Each object in the environment must be declared to be of some type.

The architecture of BABY-SIT is composed of seven major parts: *programmer/user interface*, *environment*, *background situation*, *anchoring situations*, *constraint set*, *inference engine*, and *interpreter* (Figure 1).

The interface allows interaction of the user with the system. The environment initially consists of static situation structures and their relationships. These structures can be dynamically changed and new relationships among situation types can be defined as the computation proceeds. Information conveyance among situations is made possible by defining a *part-of* relation among them. In this way, a situation s can have information about another situation s' which is part of s. The background situation contains infons which are inherited by all situation structures in the environment. However, a situation can inherit an infon from the background situation only if it does not cause a contradiction in that situation.

A situation in the environment can be realized if its parameters are anchored to objects in the real world. This is made possible by the anchoring situations which allow parameters to be anchored to objects of appropriate types—an individual, a situation, a parameter, etc. A parameter must be anchored to a unique object by an anchoring situation. On the other hand, more than one parameter may be anchored to the same object. Restrictions on parameters assure anchoring of one parameter to an object having the same qualifications as the parameter.

In addition to the *part-of* relation among situations, constraints are potent means of information conveyance between situations. They link various types of situations. Constraints may be physical laws, linguistic rules, law-like correspondences, conventions, etc. In BABY-SIT, they are realized as forward-chaining constraints or backward-chaining constraints, or both. Assertion of a new object into BABY-SIT activates the forward-chaining mechanism. Once their antecedent parts are satisfied, consequent parts of the forward-chaining constraints are asserted into BABY-SIT, unless this yields a contradiction. In case of a contradiction, the backward-chaining mechanism is activated to resolve it. The interpreter is the central authority in BABY-SIT. Anchoring of parameters, evaluation of constraints, etc., are all controlled by this part of the system.

2.3.2 Modes of Computation

A prototype of BABY-SIT is currently being developed in KEE (Knowledge Engineering Environment) (KEE^{TM} 1993) on a SPARCstation^{TM}. Some of the available modes of computation in this evolving system are described below.

2.3.2.1 Assertions

Assertion mode provides an interactive environment in which one can define objects and their types. There are nine basic types corresponding to nine primitive domains: ~IND (individuals), ~TIM (times), ~LOC (places), ~REL (relations), ~POL (polarities), ~INF (infons), ~PAR (parameters), ~SIT (situations), and ~TYP (types). For instance, if l is a place, then l is of type ~LOC, and the infon ≪*type-of*, ~LOC, l, 1≫ is a fact in the background situation. Note that type of all types is ~TYP. For example, the infons ≪*type-of*, ~TYP, ~LOC, 1≫ and ≪*type-of*, ~TYP ~TYP, 1≫ are default facts in the background situation. The syntax of the assertion mode is the same as in (Devlin 1991) (cf. Tables 1 and 2).

Suppose *fred* is an individual, *can-think* is a relation, and *s0* is a situation. Then, these objects can be declared as:

I> *fred*: ~IND

I> *can-think*: ~REL

I> *s0*: ~SIT

The definition of relations includes the *appropriateness conditions* for their argument roles. Each argument can be declared to be from one or

TABLE 1 Syntax of the Assertion Mode.

< proposition > ::=
<situation-proposition> | <parameter-type-proposition> |
<situation/object-type-proposition> | <infon-proposition>
<type-of-type-proposition> | <relation-proposition>

<situation-proposition> ::= <constant> "|=" <infonic-set>

<parameter-type-proposition> ::= <parameter> "="
{<basic-type>, <type-name>,
<restricted-parameter-type>}

<situation/object-type-proposition> ::=
<constant> ":" {<basic-type>, <type-name>
<type-abstraction>}

<infon-proposition> ::= <constant> "=" <infon>

<type-of-type-proposition> ::=
<type-name> "=" {<basic-type>, <type-abstraction>}

<relation-proposition> ::= "<" <relation> ["|" <type-specifier>
("," <type-specifier>)*] ">"
"[" (<digit>)$^+$ "]"

<type-specifier> ::= <basic-type> | <type-name> |
"{" {<basic-type>, <type-name>}
("," {<basic-type>, <type-name>})*] "}"

<type-abstraction> ::=
"[" <parameter> "|" { <constant>, <parameter>}
"|=" <infonic-set> "]"

<restricted-parameter-type> ::= <parameter> " ˆ " <infonic-set>

<basic-type> ::= "~LOC" | "~TIM" | "~IND" | "~REL" |
"~SIT" | "~INF" | "~TYP" | "~PAR" | "~POL"

<infonic-set> ::= "{" <infon> ("," <infon>)* "}" | <infon>

<infon> ::=
"≪" <relation> ("," <argument>)* ["," <polarity>] "≫"

<relation> ::= <special-relation> | <constant>

<argument> ::= <constant> | <parameter> |
<basic-type> | <type-name>

TABLE 2 Syntax of the Assertion Mode (continued).

<*polarity*> ::= "0" | "1"

<*constant*> ::= {<*digit*>, <*lower-case-letter*>}
({<*digit*>, <*lower-case-letter*>})*

<*parameter*> ::=
<*upper-case-letter*> ({<*upper-case-letter*>, <*digit*>})*

<*type-name*> ::=
"∼" <*upper-case-letter*> ({<*upper-case-letter*>, <*digit*>})*

<*lower-case-letter*> ::= "a" | "b" | ... | "z" | "-"

<*upper-case-letter*> ::= "A" | "B" | ... | "Z"

<*digit*> ::= "0" | "1" | ... | "9"

more of the primitive domains above. Consider *can-think* above. If we like it to have only one argument of type ∼IND, we can write:

I> <*can-think* | ∼IND> [1]

In order for the parameters to be anchored to objects of the appropriate type, parameters must be declared to be from only one of the primitive domains. It is also possible to put restrictions on a parameter in the environment. Suppose we want to have a parameter E that denotes any thinking individual. This can be done by asserting:

I> E = IND1 ^ ≪*can-think*, IND1, 1≫

IND1 is a default system parameter of type ∼IND. E is considered as an object of type ∼PAR such that if it is anchored to an object, say *fred*, then *fred* must be of type ∼IND and the background situation (denoted by *w*) must support the infon ≪*can-think*, *fred*, 1≫.

Parametric types are also allowed in BABY-SIT. They can be formed by obtaining a type from a parameter. This process is known as *(object-) type-abstraction.* Parametric types are of the form $[P \mid s \models I]$ where P is a parameter, s is a situation (i.e., a *grounding* situation), and I is a set of infons. The type of all thinking individuals can be defined as follows:

I> ∼HUMAN = [IND1 | *w* ⊨ ≪*can-think*, IND1, 1≫]

∼HUMAN is seen as an object of type ∼TYP and can be used as a type specifier for declaration of new objects in the environment. For instance:

I> *mary*: ∼HUMAN

yields an object, *mary*, which is of type ~IND such that the background situation supports the infon $\ll$*can-think*, *mary*, 1$\gg$.

Infons can be added into situations in BABY-SIT. The following sequence of assertions adds $\ll$*fires*, *mary*, *gun*, 1$\gg$ into *s0* (cf. Figure 3):

I> ~WEAPON = ~IND

I> *gun*: ~WEAPON

I> *fires*: ~REL

I> <*fires* | ~HUMAN, ~WEAPON>

I> *s0* $\models$ $\ll$*fires*, *mary*, *gun*, 1$\gg$

2.3.2.2 Constraints

All possible types of constraints in situation theory can be classified as either *conditional* or *unconditional.* Conditional constraints can be applied to situations that satisfy some condition while unconditional constraints can be applied to all situations.

Variables in BABY-SIT are only used in constraints and query expressions, and have scope only within the constraint or the query expression they appear. A variable can match any object appropriate for the place or the argument role it appears in. For example, given the declaration above, variables ?X and ?Y in the proposition ?S $\models$ $\ll$*fires*, ?X, ?Y, 1$\gg$ can only match objects of type ~HUMAN (i.e., of basic type ~IND where the background situation supports the fact that this object can think) and ~WEAPON (or simply ~IND), respectively.

A BABY-SIT constraint is of the form:

$antecedent_1, \ldots, antecedent_n \ \{\Leftarrow, \Rightarrow, \Leftrightarrow\}$
$\quad consequent_1, \ldots, consequent_m.$

Each $antecedent_i$, $1 \leq i \leq n$, and each $consequent_j$, $1 \leq j \leq m$, is of the form sit $\{\models, \not\models\} \ll rel, arg_1, \ldots, arg_l, pol \gg$ such that sit, rel, and each arg_k, $1 \leq k \leq l$, can either be an object of appropriate type or a variable.

Each constraint has an identifier associated with it and must belong to a group of constraints. For example, the following is a forward-chaining constraint named R6 under the constraint group GUNFIRE:

GUNFIRE:
R6:
?S1 $\models$ $\ll$*loads*, ?M, ?G, 1$\gg$ $\Rightarrow$?S1 $\models$ $\ll$*loaded*, ?G, 1$\gg$

where ?S1, ?M and ?G are variables. ?S1 can only be assigned an object of type ~SIT while ?M and ?G can have values of some type appropriate for the argument roles of *loads* and *loaded.* This constraint can apply in any situation. Hence, BABY-SIT constraints can be global. Constraints can also be situated. For example, R6 above can be rewritten to apply only in situation *sit1*:

GUNFIRE:
  R6:
    $sit1 \models \ll loads, ?M, ?G, 1 \gg \Rightarrow sit1 \models \ll loaded, ?G, 1 \gg.$

In BABY-SIT, conditional constraints come with a set of *background conditions* which must be satisfied in order for the constraint to apply. Background conditions are accepted to be true by default, unless stated otherwise. For example, to state that one hears noise upon firing a loaded gun, we can write:

GUNFIRE:
  R0:
    $?S1 \models \{\ll loaded, ?G, 1 \gg, \ll fires, ?M, ?G, 1 \gg\} \Rightarrow$
    $?S2 \models \ll hears, ?M, noise, 1 \gg$
  UNDER-CONDITIONS:
    w: $\ll exists, air, 1 \gg$.

Background conditions are, in fact, assumptions which are required to hold for constraints to be eligible for activation. R0 can become a candidate for activation only if it is the case that $w \not\models \ll exists, air, 0 \gg$, i.e., if the absence of air is not known in the background situation. Hence, background conditions provide a contextual environment for constraints (Akman and Tın 1990).

A candidate forward-chaining constraint is activated whenever its antecedent part is satisfied. All the consequences are asserted if they do not yield a contradiction in the situation into which they are asserted. New assertions may in turn activate other candidate forward-chaining constraints. Candidate backward-chaining constraints are activated either when a query is entered explicitly or is issued by the forward-chaining mechanism. In BABY-SIT, constraints between situation types as well as between infons of a situation can be easily modeled. Grouping of constraints enables one to view the world and make inferences from different perspectives. Figure 2 illustrates the axiomatization of the so-called *Yale Shooting Problem* (YSP)[1] with BABY-SIT constraints.

2.3.2.3 Querying

The query mode enables one to issue queries about situations. There are several possible actions which can be further controlled by the user:

[1] At some point in time, a person (Fred) is alive and a loaded gun, after waiting for a while, is fired at Fred. What are the results of this action? (Shoham 1988)

```
xterm
GUNFIRE R6
?S1|=<<loads,?M,?G,1>> => ?S1|=<<loaded,?G,1>>
-

GUNFIRE R5
?S1|=<<situation-p,?S1,1>>, ?S1|/=<<successor-of,?S2,?S1,1>> =>
?S3|={<<situation-p,?S3,1>>,<<time-of,?T,?S3,1>>}, ?S1|=<<successor-of,?S3,?S1,1>>
-

GUNFIRE R4
?S1|={<<alive,?F,1>>,<<successor-of,?S2,?S1,1>>} => ?S2|=<<alive,?F,1>>
-
  w: <<exists,air,1>>,
?S1: <<fires,?M,gun,0>>

GUNFIRE R3
?S1|={<<loaded,?G,1>>,<<successor-of,?S2,?S1,1>>} => ?S2|=<<loaded,?G,1>>
-
?S1: {<<fires,?M,?G,0>>,<<emptied-manually,?G,0>>}

GUNFIRE R2
?S1|={<<alive,?F,1>>,<<loaded,?G,1>>,<<fires,?M,?G,1>>,<<successor-of,?S2,?S1,1>>} =>
?S2|=<<dead,?F,1>>
-
?S1: {<<has-firing-pin,?G,1>>,<<marshmallow-bullets-in,?G,0>>}

GUNFIRE R1
?S1|={<<loaded,?G,1>>,<<fires,?M,?G,1>>,<<successor-of,?S2,?S1,1>>} =>
?S2|=<<hears,?M,noise,1>>
-
  w: <<exists,air,1>>,
?S1: {<<has-firing-pin,?G,1>>,<<marshmallow-bullets-in,?G,0>>}
```

FIGURE 2 BABY-SIT Constraints for the YSP.

- Replacing each parameter in the query expression by the corresponding individual if there is a possible anchor, either partial or full, provided by the given anchoring situation for that parameter.
- Returning solutions. (Their number is determined by the user.)
- Displaying a solution with its parameters replaced by the individuals to which they are anchored by the given anchoring situation.
- For each solution, displaying infons anchoring any parameter in the solution to an individual in the given anchoring situation.
- Displaying a trace of anchoring of parameters in each solution.

The computation upon issuing a query is done either by direct querying through situations or by the application of backward-chaining constraints. A situation, s, *supports* an infon if the infon is either explicitly asserted to hold in s, or it is supported by a situation s' which is part of s, or it can be proven to hold by application of backward-chaining constraints. The syntax of the query expressions is given in Table 3. Given *anchor1* as the anchoring situation, a query and the system's response to it are as follows (cf. Figure 3):

Q> ?U1 $\models$ {$\ll$*fires*, ?X, ?Y, 1$\gg$, $\ll$*successor-of*, ?U2, ?U1, 1$\gg$},
?U2 $\models$ {$\ll$*dead*, ?Z, 1$\gg$, $\ll$*time-of*, ?T1, ?U2, 1$\gg$}

Solution: 1

r5-1 $\models$ $\ll$*fires*, *mary*, *gun*, 1$\gg$

r5-1 $\models$ $\ll$*successor-of*, r5-2, r5-1, 1$\gg$

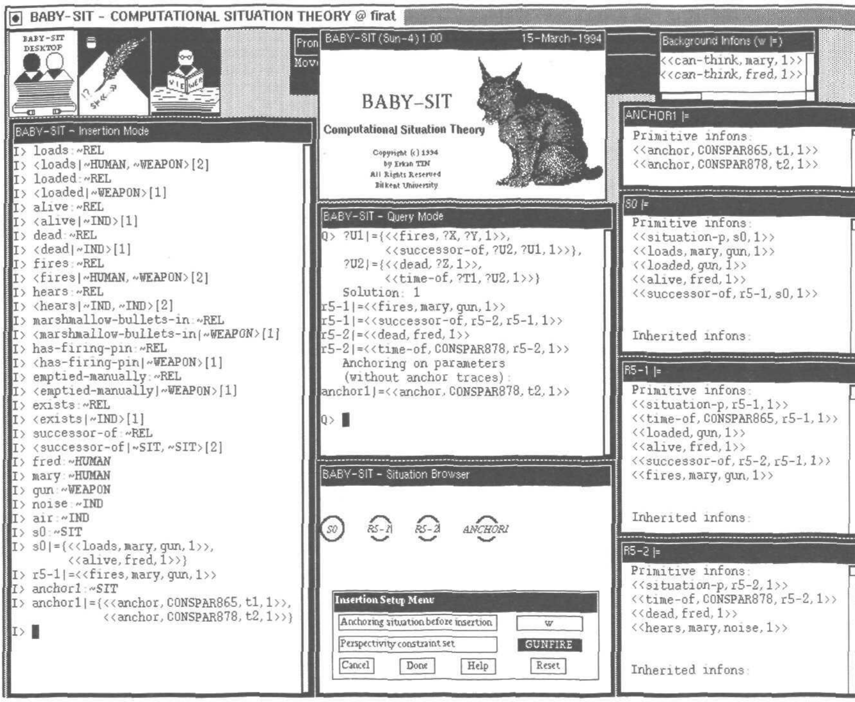


FIGURE 3 Solution of the YSP in BABY-SIT.

r5-2 $\models \ll$*dead*, fred, 1$\gg$
r5-2 $\models \ll$*time-of*, CONSPAR878, r5-2, 1$\gg$
Anchoring on parameters (without anchor traces):
anchor1 $\models \ll$*anchor*, CONSPAR878, t2, 1$\gg$.

In addition to query operations, a special operation, *oracle*, is allowed in the query mode. An *oracle* is defined over an object and a set of infons (*set of issues*) (Devlin 1991). The oracle of an object enables one to chronologically view the information about that object from a particular perspective provided by the given set of infons. One may consider oracles as 'histories' of specific objects. Given an object and a set of issues, BABY-SIT anchors all parameters in this set of issues and collects all infons supported by the situations in the system under a specific situation, thus forming a 'minimal' situation which supports all parameter-free infons in the set of issues.

2.4 Concluding Remarks

BABY-SIT accommodates the basic features of situation theory and, compared with existing approaches, enhances these features (cf. Table 4). Situations are viewed at an abstract level. This means that situations are sets of

TABLE 3 Syntax of the Query Mode

```
< query > ::= <situation-query> | <oracle-query>

<situation-query> ::=
    <situation> {"|=", "|=/="} <query-infonic-set>
    ("," <situation> {"|=", "|/="} <query-infonic-set>)*

<oracle-query> ::=
    <constant> "=" "@" "(" <constant> ")" [<issue-set>]

<situation> ::= <constant> | <query-variable>

<issue-set> ::= "{" <issue-infon> ("," <issue-infon>)* "}"

<query-infonic-set> ::=
    "{" <query-infon> ("," <query-infon>)* "}" | <query-infon>

<query-infon> ::=
    "≪" {<relation>, <query-variable>}
    ("," {<argument>, <query-variable>})* "," <polarity> "≫"

<issue-infon> ::=
    "≪" <relation> ("," <argument>)* "," <polarity> "≫"

<query-variable> ::= "?" <parameter>
```

parametric infons, but they may be non-well-founded. Parameters are place holders, hence they can be anchored to unique individuals in an anchoring situation. A situation can be realized if its parameters are anchored, either partially or fully, by an anchoring situation. Each relation has 'appropriateness conditions' which determine the type of its arguments. Situations (and hence infons they support) may have spatio-temporal dimensions. A hierarchy of situations can be defined both statically and dynamically. A built-in structure allows one situation to have information about another which is part of the former. Grouping of situations provides a computational context. Partial nature of the situations facilitates computation with incomplete information. Constraints can be violated. This aspect is built directly into the computational mechanism: a constraint can be applied to a situation only if it does not lead to an incoherence.

With these features, BABY-SIT provides a programming environment incorporating situation-theoretic constructs for various domains of application including artificial intelligence and computational linguistics. A pre-

TABLE 4 Tableau Comparison of Existing Approaches.

Constraint Type	PROSIT	ASTL	BABY-SIT
Nomic	√	√	√
Necessary	√	√	√
Conventional	–	–	?
Conditional	–	–	√
Situated	√	–	–

Computation	PROSIT	ASTL	BABY-SIT
Unification	√	√	√
Type-theoretic	–	–	√
Coherence	–	–	√
Forward-chaining	√	–	√
Backward-chaining	√	√	√
Bidirectional-chaining	√	–	√

Miscellaneous Features	PROSIT	ASTL	BABY-SIT
Circularity	√	√	√
Partiality	√	√	√
Parameters	?	?	√
Abstraction	?	?	√
Anchoring	?	?	√
Appropriateness conditions	–	–	√
Information nesting	√	√	√
Set operations	√	–	–
Oracles	–	–	?

LEGEND: √: available, – : not available, and ? : partially/conceptually available.

liminary study towards employing BABY-SIT in the resolution of pronominal anaphora has been recently initiated (Tın and Akman 1994c).

References

Akman, V., and E. Tın. 1990. What is in a Context? In *Proceedings of the 1990 Bilkent International Conference on New Trends in Communication, Control, and Signal Processing*, ed. E. Arıkan. 1670–1676. Amsterdam, Holland: Elsevier.

Barwise, J., and J. Etchemendy. 1987. *The Liar: An Essay on Truth and Circularity.* New York: Oxford University Press.

Barwise, J., and J. Perry. 1983. *Situations and Attitudes.* Cambridge, Massachusetts: MIT Press.

Black, A. W. 1993. *An Approach to Computational Situation Semantics.* Doctoral dissertation, Department of Artificial Intelligence, University of Edinburgh, Edinburgh, U.K.

Devlin, K. 1991. *Logic and Information.* Cambridge, U.K.: Cambridge University Press.

Dretske, F. 1981. *Knowledge and the Flow of Information.* Cambridge, Massachusetts: MIT Press.

Fenstad, J. E., P.-K. Halvorsen, T. Langholm, and J. van Benthem. 1987. *Situations, Language, and Logic.* Dordrecht, Holland: Reidel.

Israel, D., and J. Perry. 1990. What is Information? In *Information, Language, and Cognition*, ed. P. P. Hanson. 1–28. Vancouver, Canada: The University of British Columbia Press.

KEE^{TM}. 1993. (Knowledge Engineering Environment) Software Development System, Version 4.1. IntelliCorp, Inc., Mountain View, California.

Nakashima, H., H. Suzuki, P.-K. Halvorsen, and S. Peters. 1988. Towards a Computational Interpretation of Situation Theory. In *Proceedings of the International Conference on Fifth Generation Computer Systems*, 489–498. Institute for New Generation Computer Technology, Tokyo, Japan.

Schütze, H. 1991. The PROSIT Language, Version 0.4 CSLI Memo, Center for the Study of Language and Information, Stanford University, Stanford, California.

Shoham, Y. 1988. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence.* Cambridge, Massachusetts: MIT Press.

Tin, E., and V. Akman. 1994a. BABY-SIT: Towards a Situation-Theoretic Computational Environment. In *Current Issues in Mathematical Linguistics*, ed. C. Martín-Vide. 299–308. Amsterdam, Holland: North-Holland.

Tin, E., and V. Akman. 1994b. Computational Situation Theory. *ACM Sigart Bulletin* 5(4):4–17.

Tin, E., and V. Akman. 1994c Situated Processing of Pronominal Anaphora. In *Tagungsband KONVENS '94 Verarbeitung Natürlicher Sprache*, ed. H. Trost, 369–378. Vienna, Austria. Informatik Xpress.

3

Reasoning with Diagram Sequences

MICHAEL ANDERSON

Abstract

A diagram sequence can be defined as a meta-diagram composed of a number of sub-diagrams arranged in an order that incorporates some manner of forward moving time. Humans possess a highly developed ability to reason with diagram sequences. Endowing a computer with such an ability could be of great benefit in terms of both human-computer interaction and computational efficiency. To facilitate formal reasoning about diagram sequences, a system of logic is proposed where diagrams themselves are treated as first-class objects. When applied to two sub-diagrams in a diagram sequence, Boolean operators can be used to help parse them and for subsequent reasoning.

Introduction

Humans possess a highly developed ability to reason with visual information such as diagrams. It has been shown that endowing a computer with such an ability could be of great benefit in terms of both human-computer interaction and computational efficiency through explicit representation (Larkin and Simon 1987). To date, research in diagrammatic reasoning has dealt with *intra*-diagrammatic reasoning almost to the exclusion of *inter*-diagrammatic reasoning that incorporates change over time (Narayanan 1992, 1993, and Chandrasekana et al. 1993). One can argue that something might be learned about the former through the investigation of the latter.

A *diagram sequence* can be defined as a meta-diagram composed of a number of sub-diagrams arranged in an order that incorporates some manner of forward moving time. Often, each of the sub-diagrams can be con-

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

sidered a discrete snapshot of some continuous phenomenon described by the sequence as a whole. Further, diagram sequences frequently appear in conjunction with a textual description of pictured phenomenon. Examples of such diagram sequences include chess notation, chordal musical notation, assembly instructions, and instructions for a product's use.

We define a logic for diagrams in Section 1, provide a number of example uses of our logic in Section 2, discuss related work in Section 3, and, finally, offer our conclusions in Section 4.

3.1 Diagrammatic Logic

Formal reasoning about diagram sequences can best proceed given an appropriate representation. To this end, a straightforward system of logic is proposed where diagrams themselves are treated as first-class objects. It is important to note that what is being proposed is an attempt to use diagrams in a formal way that has been here-to-fore largely ignored. Although Boolean operators have been used in a similar context in the past, they must be considered as simply pixel manipulators since they have not been used to the end of inferring meaning from what they operate upon. Such an endeavor requires a clear statement of syntax and semantics to provide a solid foundation for future work. The logical and non-logical symbols, syntax, and semantics of such a logic follow. A number of theorems are then postulated.

Logical Symbols

Logical symbols used include standard symbols used to represent NOT, AND, OR, and XOR:$\neg$,$\wedge$,$\vee$,$\oplus$ and respectively. Parentheses can be used to override the standard order of evaluation.

Non-logical Symbols

Non-logical symbols are comprised of *pixel constants* and *pixmat constants.*

Pixel constants are symbolized by lower-case italicized letters (p, q, r, etc.) denoting a single pixel. Pixel constants can be subscripted by subscripts (i, j, m, n, etc.) denoting the row and column the pixel occupies in a given pixmat.

Pixmat constants are symbolized by

$$\begin{matrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n} \end{matrix}$$

denoting *pixel mat*rices of m rows and n columns where $m \geq 1$ and $n \geq 1$.

Syntax

A sentence in the diagrammatic logic is defined as follows:

Definition 1 ϕ is a sentence if and only if

1. ϕ is a pixmat or pixel constant (i.e. pixmat where $m = n = 1$)
2. ϕ has one of the following forms (where α and β are sentences that evaluate to pixmats with equivalent m's and n's): $\neg\alpha$, $\alpha \vee \beta$, $\alpha \wedge \beta$, $\alpha \oplus \beta$

Semantics

The diagrammatic logic defines its atomic values to be pixels with a semantic domain of $\{\bullet, \circ\}$, denoting on and off. The semantics of the logical symbols within this domain are defined as follows:

Definition 2 Let P be a set of pixel constants, $\{p, q, ...\}$, and let I be an interpretation over P. Then

i. $I(\neg p) = \bullet$ if $I(p) = \circ$ otherwise $I(\neg p) = \circ$

ii. $I(p \vee q) = \bullet$ if either $I(p) = \bullet$ or $I(q) = \bullet$; otherwise $I(p \vee q) = \circ$

iii. $I(p \wedge q) = \bullet$ if both $I(p) = \bullet$ and $I(q) = \bullet$; otherwise $I(p \wedge q) = \circ$

iv. $I(p \oplus q) = \bullet$ if either $I(p) = \bullet$ or $I(q) = \bullet$ but not both; otherwise $I(p \oplus q) = \circ$

v.
$$I\begin{pmatrix} & p_{1,1} & \cdots & p_{1,n} \\ \neg & \vdots & \vdots & \vdots \\ & p_{m,1} & \cdots & p_{m,n} \end{pmatrix} = \begin{matrix} \neg p_{1,1} & \cdots & \neg p_{1,n} \\ \vdots & \vdots & \vdots \\ \neg p_{m,1} & \cdots & \neg p_{m,n} \end{matrix}$$

vi.
$$I\begin{pmatrix} p_{1,1} & \cdots & p_{1,n} & & q_{1,1} & \cdots & q_{1,n} \\ \vdots & \vdots & \vdots & \vee & \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n} & & q_{m,1} & \cdots & q_{m,n} \end{pmatrix} = \begin{matrix} p_{1,1} \vee q_{1,1} & \cdots & p_{1,n} \vee q_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} \vee q_{m,1} & \cdots & p_{m,n} \vee q_{m,n} \end{matrix}$$

vii.
$$I\begin{pmatrix} p_{1,1} & \cdots & p_{1,n} & & q_{1,1} & \cdots & q_{1,n} \\ \vdots & \vdots & \vdots & \wedge & \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n} & & q_{m,1} & \cdots & q_{m,n} \end{pmatrix} = \begin{matrix} p_{1,1} \wedge q_{1,1} & \cdots & p_{1,n} \wedge q_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} \wedge q_{m,1} & \cdots & p_{m,n} \wedge q_{m,n} \end{matrix}$$

viii.
$$I\begin{pmatrix} p_{1,1} & \cdots & p_{1,n} & & q_{1,1} & \cdots & q_{1,n} \\ \vdots & \vdots & \vdots & \oplus & \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n} & & q_{m,1} & \cdots & q_{m,n} \end{pmatrix} = \begin{matrix} p_{1,1} \oplus q_{1,1} & \cdots & p_{1,n} \oplus q_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} \oplus q_{m,1} & \cdots & p_{m,n} \oplus q_{m,n} \end{matrix}$$

Less formally, Boolean (*pix*ean?) operators can be applied to pixmats by applying unary operators to each pixel of a pixmat and binary operators to correspondingly indexed pair of pixels within each of two pixmats of equal dimensions. A negated ($\neg$) pixmat will yield another pixmat with all the original pixel values reversed. When two pixmats have equal m's and n's (often the case in many diagram sequences), the binary operators OR ($\vee$), AND ($\wedge$), and XOR ($\oplus$) can be applied to them. ORing two such pixmats produces a new pixmat that incorporates all $\bullet$ pixels from each pixmat. ANDing two such pixmats produces a new pixmat that has $\bullet$ pixels only where corresponding pixels in both are $\bullet$. XORing two such pixmats produces a new pixmat that has $\bullet$ pixels only where exactly one of corresponding pixels in both are $\bullet$. These notions are made formal in the following theorems.

Theorems

The *Pixel Commonality* theorem states that ANDing two pixmats of equal dimensions produces a new pixmat that is comprised only of those pixels that are in both. For example,

$$\begin{matrix}\bullet & \circ \\ \bullet & \circ\end{matrix} \wedge \begin{matrix}\circ & \bullet \\ \bullet & \circ\end{matrix} \Rightarrow \begin{matrix}\circ & \circ \\ \bullet & \circ\end{matrix}.$$

Applying this theorem to two sub-diagrams in a diagram sequence will yield a new diagram that is comprised of only the pixels that are common to both. This can be considered as the background that is unchanged between two sub-diagrams.

Theorem 1 *Pixel Commonality*

$$\begin{matrix}p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n}\end{matrix} \wedge \begin{matrix}q_{1,1} & \cdots & q_{1,n} \\ \vdots & \vdots & \vdots \\ q_{m,1} & \cdots & q_{m,n}\end{matrix} \Rightarrow \begin{matrix}r_{1,1} & \cdots & r_{1,n} \\ \vdots & \vdots & \vdots \\ r_{m,1} & \cdots & r_{m,n}\end{matrix}$$

where $r_{i,j} = \bullet$ *if and only if* $p_{i,j} = \bullet$ *and* $q_{i,j} = \bullet$ *otherwise* $r_{i,j} = \circ$ *, for* $i = 1$ *to* m *and* $j = 1$ *to* n.

Proof. $r_{i,j} \equiv p_{i,j} \wedge q_{i,j}$ by Definition vii. □

The *Pixel Merging* theorem states that ORing two pixmats of equal dimensions produces a new pixmat that is comprised of all $\bullet$ pixels in both. For example,

$$\begin{matrix}\bullet & \circ \\ \bullet & \circ\end{matrix} \vee \begin{matrix}\circ & \bullet \\ \bullet & \circ\end{matrix} \Rightarrow \begin{matrix}\bullet & \bullet \\ \bullet & \circ\end{matrix}$$

Applying this theorem to two sub-diagrams in a diagram sequence will yield a new diagram that merges all pixels in both. This can be used to generate new diagrams comprised of others.

Theorem 2 *Pixel Merging*

$$\begin{array}{ccc} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n} \end{array} \vee \begin{array}{ccc} q_{1,1} & \cdots & q_{1,n} \\ \vdots & \vdots & \vdots \\ q_{m,1} & \cdots & q_{m,n} \end{array} \Rightarrow \begin{array}{ccc} r_{1,1} & \cdots & r_{1,n} \\ \vdots & \vdots & \vdots \\ r_{m,1} & \cdots & r_{m,n} \end{array}$$

where $r_{i,j} = \bullet$ *if and only if* $p_{i,j} = \bullet$ *or* $q_{i,j} = \bullet$ *otherwise* $r_{i,j} = \circ$, *for* $i = 1$ *to* m *and* $j = 1$ *to* n.

Proof. $r_{i,j} \equiv p_{i,j} \vee q_{i,j}$ by Definition vi. □

The *Pixel Difference* theorem states that XORing two pixmats of equal dimensions produces a new pixmat that is comprised of pixels that are • in exactly one or the other. For example,

$$\begin{array}{cc} \bullet & \circ \\ \bullet & \circ \end{array} \oplus \begin{array}{cc} \circ & \bullet \\ \bullet & \circ \end{array} \Rightarrow \begin{array}{cc} \bullet & \bullet \\ \circ & \circ \end{array}.$$

Applying this theorem to two sub-diagrams in a diagram sequence will yield a new diagram that consists of only those pixels that are different in each. This new diagram can be considered the combined differences of the two sub-diagrams.

Theorem 3 *Pixel Difference*

$$\begin{array}{ccc} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n} \end{array} \oplus \begin{array}{ccc} q_{1,1} & \cdots & q_{1,n} \\ \vdots & \vdots & \vdots \\ q_{m,1} & \cdots & q_{m,n} \end{array} \Rightarrow \begin{array}{ccc} r_{1,1} & \cdots & r_{1,n} \\ \vdots & \vdots & \vdots \\ r_{m,1} & \cdots & r_{m,n} \end{array}$$

where $r_{i,j} = \bullet$ *if and only if either* $p_{i,j} = \bullet$ *or* $q_{i,j} = \bullet$ *but not both otherwise* $r_{i,j} = \circ$, *for* $i = 1$ *to* m *and* $j = 1$ *to* n.

Proof. $r_{i,j} \equiv p_{i,j} \oplus q_{i,j}$ by Definition viii. □

The *Pixel Introduction* theorem states that, given two pixmats of equal dimensions, negating the first and ANDing it with the second will produce a new pixmat comprised of only the pixels introduced in the second. For example,

$$\neg \begin{array}{cc} \bullet & \circ \\ \bullet & \circ \end{array} \wedge \begin{array}{cc} \circ & \bullet \\ \bullet & \circ \end{array} \Rightarrow \begin{array}{cc} \circ & \bullet \\ \circ & \circ \end{array}.$$

Applying this theorem to two sub-diagrams in a diagram sequence will yield a new diagram that consists of only what was added in the second. This new diagram can be considered as representing what was introduced over time.

Theorem 4 *Pixel Introduction*

$$\neg \begin{array}{ccc} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n} \end{array} \wedge \begin{array}{ccc} q_{1,1} & \cdots & q_{1,n} \\ \vdots & \vdots & \vdots \\ q_{m,1} & \cdots & q_{m,n} \end{array} \Rightarrow \begin{array}{ccc} r_{1,1} & \cdots & r_{1,n} \\ \vdots & \vdots & \vdots \\ r_{m,1} & \cdots & r_{m,n} \end{array}$$

where $r_{i,j} = \bullet$ *if and only if* $p_{i,j} = \circ$ *and* $q_{i,j} = \bullet$ *otherwise* $r_{i,j} = \circ$, *for* $i = 1$ *to* m *and* $j = 1$ *to* n.

Proof.

$$\neg \begin{matrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n} \end{matrix} \wedge \begin{matrix} q_{1,1} & \cdots & q_{1,n} \\ \vdots & \vdots & \vdots \\ q_{m,1} & \cdots & q_{m,n} \end{matrix}$$

by assumption

$$\begin{matrix} \neg p_{1,1} & \cdots & \neg p_{1,n} \\ \vdots & \vdots & \vdots \\ \neg p_{m,1} & \cdots & \neg p_{m,n} \end{matrix} \wedge \begin{matrix} q_{1,1} & \cdots & q_{1,n} \\ \vdots & \vdots & \vdots \\ q_{m,1} & \cdots & q_{m,n} \end{matrix}$$

by Definition v $\therefore r_{i,j} \equiv \neg p_{i,j} \wedge q_{i,j}$ by Definition vii. $\square$

The *Pixel Removal Theorem* states that, given two pixmats of equal dimensions, negating the second and ANDing it with the first will produce a new pixmat comprised of only the pixels removed in the second. For example,

$$\begin{matrix} \bullet & \circ \\ \bullet & \circ \end{matrix} \wedge \neg \begin{matrix} \circ & \bullet \\ \bullet & \circ \end{matrix} \Rightarrow \begin{matrix} \bullet & \circ \\ \circ & \circ \end{matrix}.$$

Applying this theorem to two sub-diagrams in a diagram sequence will yield a new diagram that consists of only what was removed in the second. This new diagram can be considered as representing what was removed over time.

Theorem 5 *Pixel Removal*

$$\begin{matrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n} \end{matrix} \wedge \neg \begin{matrix} q_{1,1} & \cdots & q_{1,n} \\ \vdots & \vdots & \vdots \\ q_{m,1} & \cdots & q_{m,n} \end{matrix} \Rightarrow \begin{matrix} r_{1,1} & \cdots & r_{1,n} \\ \vdots & \vdots & \vdots \\ r_{m,1} & \cdots & r_{m,n} \end{matrix}$$

where $r_{i,j} = \bullet$ *if and only if* $p_{i,j} = \bullet$ *and* $q_{i,j} = \circ$ *otherwise* $r_{i,j} = \circ$, *for* $i = 1$ *to* m *and* $j = 1$ *to* n.

Proof.

$$\begin{matrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n} \end{matrix} \wedge \neg \begin{matrix} q_{1,1} & \cdots & q_{1,n} \\ \vdots & \vdots & \vdots \\ q_{m,1} & \cdots & q_{m,n} \end{matrix}$$

by assumption

$$\begin{matrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \vdots \\ p_{m,1} & \cdots & p_{m,n} \end{matrix} \wedge \begin{matrix} \neg q_{1,1} & \cdots & \neg q_{1,n} \\ \vdots & \vdots & \vdots \\ \neg q_{m,1} & \cdots & \neg q_{m,n} \end{matrix}$$

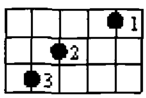
by Definition v $\therefore r_{i,j} \equiv p_{i,j} \wedge \neg q_{i,j}$ by Definition vii. $\square$

3.2 Diagram Parsing

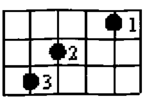
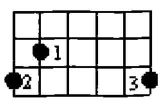
When applied to two sub-diagrams in a diagram sequence, such theorems can be used to help parse them. Diagram parsing and subsequent inferencing can be illustrated through three example diagram sequence domains: *guitar chord notation, chess notation,* and *product use instructions.* These domains differ along various dimensions including granularity, regularity, and ambiguity.

Guitar Chord Notation

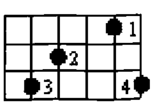
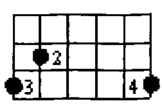
Guitar chord notation is a well-developed symbol system of fine granularity and unambiguous syntax although somewhat ambiguous semantics. Syntactically, vertical lines represent the *strings* of a guitar whereas horizontal lines represent its *frets.* A dot on a string represents where some finger is placed to produce a desired pitch. Semantically, a *fingering* is a specification of exactly which of four fingers to use to realize the dots of the diagram. For example, given that numbers 1 through 4 represent the index finger to the little finger, the following is a chord diagram complete with a fingering:



Chord diagrams are superior to standard musical notation for inferring fingering information since the fingerboard positioning of the chord is explicitly shown on the diagram but must be inferred from standard musical notation. Even so, semantic ambiguity arises in guitar chord diagrams because 1) fingerings are often not specified and 2) there exists a one-to-many mapping between the dots and possible fingerings. A given chord can sometimes be fingered many ways with the preferred way often being context dependent. That is, the preferred fingering of a chord will often depend on one or both of the chords preceding and following it in the diagram sequence. For example, when

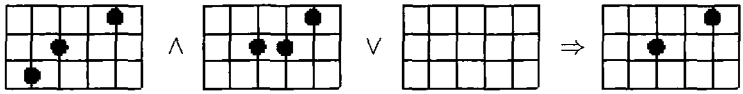
 precedes 

there is no dot in common between them and, therefore, the fingering for the second chord defaults to its least demanding state as shown. If, however,

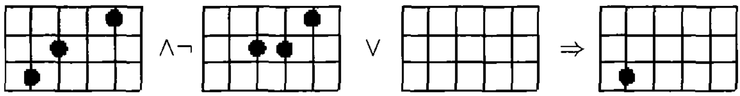
 precedes 

there is a dot in common and the fingering for the second chord attempts to conserve finger movement by leaving the fourth finger in place as shown.

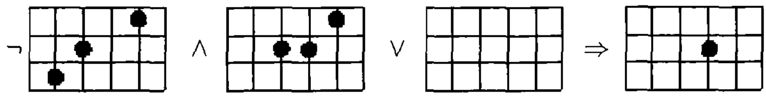
Such diagrams can be parsed using the previously postulated theorems:



by Pixel Commonality & Merging

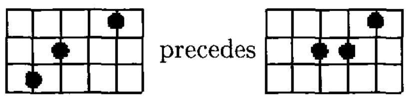


by Pixel Removal & Merging



by Pixel Introduction & Merging

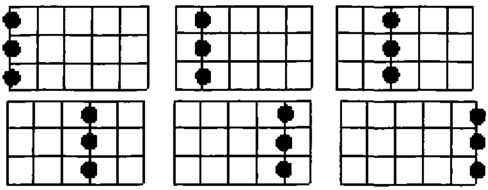
The results of these theorems can be used in subsequent reasoning about chord diagrams. For example, given a sequence in which



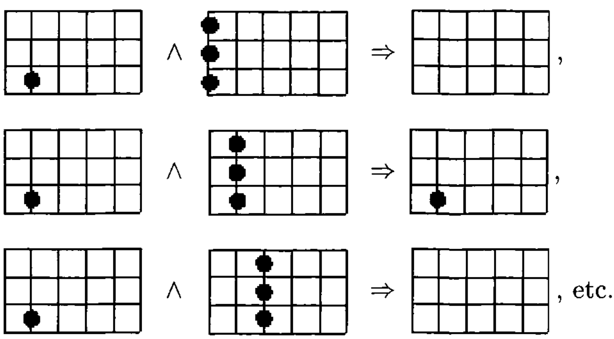
a fingering for the right hand diagram can be inferred from a fingering for the left via diagrams generated by logical operations in concert with a few simple rules: 1) whenever possible keep fingers in the same position, 2) use next numerically available finger, and 3) only fingers not currently in use are available.

Since there can only be one finger on any given string, we can represent a fingering for a given chord by a *fingering vector*, $[s_6, s_5, s_4, s_3, s_2, s_1]$, where each s_i is a finger number 0 through 4 signifying which (if any) finger is to be placed on the string i. (The strings on a guitar are numbered from lowest pitch to highest pitch as 6 through 1.) The fingering that will be used for the first chord in this example will be represented as $[0, 3, 2, 0, 1, 0]$. A list of available fingers can be represented by an *available finger set* that contains the numbers of all fingers not currently in use by a chord. This can be generated for any given chord by inspecting its fingering vector. Thus, the fingers not in use by the first chord in the example is represented as $\{4\}$.

The first step to inferring a fingering for the second chord is to update the available finger set with newly available fingers. These can be found by inspecting the results of applying Pixel Removal to the chord diagrams. Any finger that was used to realize a dot that was removed from the first chord is now available. To accomplish this inspection, six *inspection diagrams*, id_i, are defined, each associated with one string in the diagram:

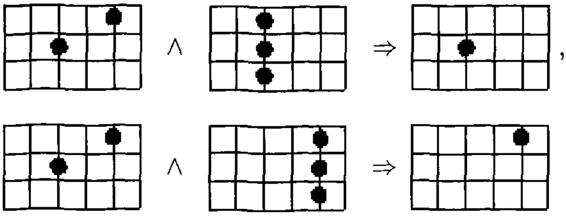


When each of these are individually ANDed with a given diagram, a background grid devoid of dots will result whenever there is no dot on the string associated with the inspection diagram in the diagram under inspection. Further, whenever there does exist a dot on the currently inspected string, such ANDing will infer a diagram comprised of the background grid with that dot in place by Pixel Commonality. In the example,



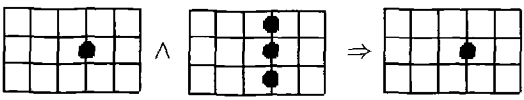
Since the only inference that produces anything other than the background grid is the one involving the fifth string's inspection diagram, the finger that was used on s_5 in the fingering vector (namely, finger number three) is now available. The available finger set is then updated to $\{3, 4\}$.

A similar approach is taken to find which fingers (if any) should remain in the positions they occupied in the first chord diagram. Each inspection diagram is ANDed, in turn, to the results of applying the Pixel Commonality theorem. The inferences that do not produce an empty background grid are exactly those made with inspection diagrams associated with the strings that should maintain the same fingering. These fingers are then transferred to the new fingering under development. In the example,



therefore, since the first and second fingers were used on the second and fourth strings, respectively, the partial fingering for the next chord that keeps these the fingers the same is $[0, 0, 2, 0, 1, 0]$ given a fingering vector that is initialized as containing all zeros

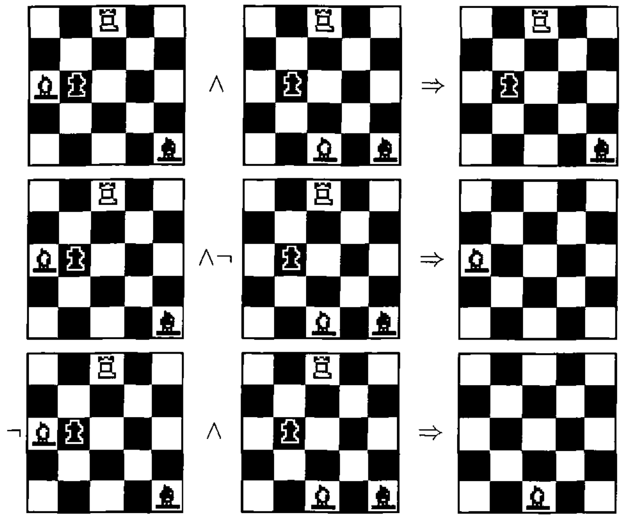
Next, the result of Pixel Introduction is used to find which strings require new fingers (if any). Again, the inspection diagrams are used to find which diagrams have dots in this result. Those that do need to have fingers placed on their associated strings. In the example,



is the only inference that results in more than a background grid, so a new finger is needed on the third string. This finger is retrieved from the available finger set via a *minimum function* that returns the smallest finger number in the set. In the example, the minimum number in the available finger set is 3. This finger number is then placed in the vector at the place corresponding to the string that needs a finger, s_3. In the example, the final fingering vector for the second chord diagram is $[0, 0, 2, 3, 1, 0]$.

Chess Notation

Chess notation is a well-developed symbol system of medium granularity and unambiguous syntax and semantics. The postulated theorems can be applied to this type of diagram as well. For example:



by Pixel Commonality

by Pixel Removal

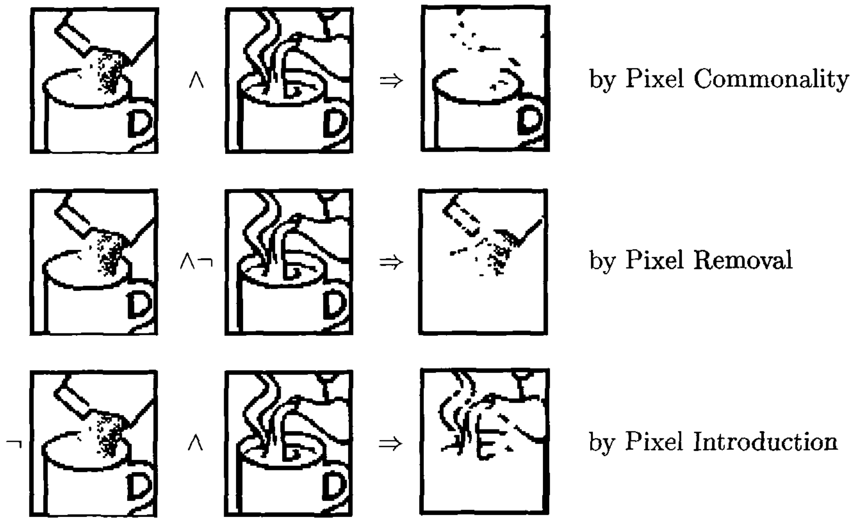
by Pixel Introduction

(A background board has been merged with the result of each of the previous theorem applications.)

Further enhancing the logic to define *ranks, rows*, and *diagonals*, the results of applying logical operations to a sequence of two chess diagrams can be used, combined with rules of movement, to verify the validity of the chess move represented by the two. Comparison of removal and introduction results could help confirm that movement has been along appropriate lines, inspection of removal results help could confirm that captured pieces belong to the opposition, and commonality results could be used to verify the absence of intervening pieces when required.

Product Use Instructions

Product use instructions are generally not based on well-developed symbol systems. There is no set syntax, although there have been some attempts to classify categories of symbols that occur across domains (Beiger and Glock 1985). It is large grained in that objects that are portrayed are often comparitively complex. Further, semantic content varies from domain to domain. Even within such a daunting framework, some use can often be made of applications of the postulated theorems. For example, given two contiguous sub-diagrams from a diagram sequence that attempts to give instructions for making cocoa, the following results are obtained from theorem applications:



(A background frame has be merged with the result of each of the previous theorem applications.)

Product use diagrams might be so parsed into instantiations of objects referred to in the textual instructions that often accompany them. Recognition of graphical transformations of these referenced entities might facilitate the integration of textual and diagrammatic information and, therefore, subsequent reasoning.

3.3 Related Work

As previously stated, little work has been done with diagram sequences per se. One notable exception is the work done by Bieger and Glock (1985 and 1986). Beside attempting a taxonomy of categories of information presented in what they term *picture-text instructions*, they performed rigorous experimentation with actual subjects and monitored their use of such instructions to the end of identifying the most critical of such categories. The direction of their work was not towards automating diagrammatic reasoning but towards understanding human use of such information as is the work in Willows and Houghton 1987.

Furnas 1992 postulates a logic that deals with diagrams via BITPICT rule mappings that can be used to transform one diagram into another and, therefore, allows reasoning from diagrams to diagrams. Interesting as this reasoning is, these explicit rule mappings can be subsumed by the definitions of the more general logic currently proposed. They are simply specific combinations of pixel commonality, removal, and introduction theorem applications. Further, Furnas's work does not attempt to reason *about* diagrams in sequences but, rather, its crux is the *generation* of sequences of diagrams to accomplish some reasoning goal pertaining to a single diagram.

3.4 Conclusion

In conclusion, it has been shown that a logic that grants diagrams the status formerly reserved for sentential representations can help a system to reason with sequences of diagrammatic representations directly. Diagram sequences have not been the topic of diagrammatic reasoning in general and it is expected that investigation of such will lend a fresh perspective to the field as a whole.

Future work will attempt to extend this logic, employ it as a foundation for diagrammatic man-machine interfacing, and explore how such a logic might enhance computational efficiency in various domains.

References

Bieger, G., and M. Glock. 1985. The Information Content of Picture-Text Instructions. *The Journal of Experimental Education* 53(2):68–76.

Bieger, G., and M. Glock. 1986. Comprehending Spatial and Contextual Information in Picture-Text Instructions. *The Journal of Experimental Education* 54(4):181–188.

Chandrasekana, B., N. Narayanan, and Y. Iwasaki. 1993. Reasoning with Diagrammatic Representations. *AI Magazine* 14(2).

Furnas, G. 1992. Reasoning with Diagrams Only. In *Working Notes of AAAI Spring Symposium on Reasoning with Diagrammatic Representations*, ed. N. Narayanan.

Larkin, J., and H. Simon. 1987. Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science* 11:65–99.

Narayanan, N. (ed.). 1992. *Working Notes of AAAI Spring Symposium on Reasoning with Diagrammatic Representations.*

Narayanan, N. 1993. The Imagery Debate Revisited. *Computational Intelligence* 9(4).

Willows, D., and H. Houghton. 1987. *The Psychology of Illustration: Volume 1.* Springer-Verlag.

4

Information Flow and the Lambek Calculus

JON BARWISE, DOV GABBAY AND CHRYSAFIS HARTONAS

Abstract

This paper is an investigation into the logic of information flow. The basic perspective is that logic flows in virtue of constraints (as in Barwise and Perry 1983), and that constraints classify channels connecting particulars (as in Barwise and Seligman 1993). In this paper we explore some logics intended to model reasoning in the case of idealized information flow, that is, where the constraints involved are exceptionless. We look at this as a step toward the far more challenging task of understanding the logic of imperfect information flow, that is where the constraints admit of exceptional connections.

4.1 Modeling Information Networks

Over the past decade, *information* has emerged as an important topic in the study of logic, language, and computation. This paper contributes to this line of work. We formulate a notion of *information network* that covers many important examples from the literature of information-theoretic structures. Information networks have two kinds of items. Following Barwise 1993 we call these items: *sites* of information, and *channels* between sites. We present various logical calculi intended to model perfect reasoning about the flow of information through an information network.

Definition 1 An *information network* is a structure of the form $\mathcal{N} = \langle S, C, \rightsquigarrow, \circ \rangle$ where S is a set of objects called *sites*, C is a set of objects called *channels*, $\rightsquigarrow$ is a relation on $S \times C \times S$, and $\circ$ is an associative binary operation on C. The signaling relation $s \overset{c}{\rightsquigarrow} t$ is read c is a *channel* from *source* s to *target* t. A *connection* for the channel c is a pair $\langle s, t \rangle$ such that

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

$s \overset{c}{\rightsquigarrow} t$. The channel $a \circ b$ is called the composition of a and b. The signaling relation and composition operation are required to satisfy the following condition:

- for all channels a and b,

$$\forall s,t\ [\ s \overset{a \circ b}{\rightsquigarrow} t \quad \text{iff} \quad \exists r\ (\ s \overset{a}{\rightsquigarrow} r \text{ and } r \overset{b}{\rightsquigarrow} t\ )\ ].$$

The general set-up allows for the possibility that S and C could have elements in common, or even be the very same set. If $C \subseteq S$ we call the network *homogeneous.* Otherwise it is called *heterogeneous.*

Example 1 In Barwise and Seligman 1993, it is argued that every real world information system, be it computer, the internet, language, a proof system, or whatever, consists of components connected together, the connections allowing one component to carry information about other components. If this is correct, then every such system can be modeled as an information network in the sense of this paper.

The remaining examples are intended to show how the notion of an information network subsumes a number of other mathematical structures used to model information flow.

Example 2 Let w be "the world" and, let $S = C = \{w\}$, let $w \overset{w}{\rightsquigarrow} w$ and let $w \circ w = w$. With this network, our logic will just reduce to classical propositional logic for the connectives $\wedge$ and $\rightarrow$, except that we will have two copies of each.

Example 3 Let S consist of some set of "worlds" and let $C = \{\leq\}$ be some transitive accessibility relation on S, with $s \overset{\leq}{\rightsquigarrow} t$ iff $s \leq t$ and $\leq \circ \leq = \leq$.

Example 4 Generalizing Example 3, let S consist of any set, C the set of binary relations on S, with $s \overset{c}{\rightsquigarrow} t$ iff $\langle s,t \rangle \in c$. Let $\circ$ be composition of relations.

Example 5 Let S be the set of hereditarily finite sets on some set A and let C consist of those elements of S which are binary relations, i.e., finite sets of ordered pairs of elements of $A \cup S$. Again define $s \overset{c}{\rightsquigarrow} t$ iff $\langle s,t \rangle \in c$. Notice that this example, unlike the previous two, has channels which are also sites.

Example 6 For another example where channels are also sites, let S and C both be the set of natural numbers, with $s \overset{c}{\rightsquigarrow} t$ iff s is in the domain of the unary recursive function φ^1_c whose Gödel number is c and $\varphi^1_c(s) = t$.

The logic we are going to explore turns out to be closely related to the Lambek Calculus [Lambek 1958]. To see why, note that any associative operation $\circ$ on a set C gives rise to an information network in a natural way. Let $S = C$ and define $s \overset{c}{\rightsquigarrow} t$ iff $s \circ c = t$. Then it is routine to see that $\mathcal{N} = \langle S, C, \rightsquigarrow, \circ \rangle$ is an information network. We will call a network of this kind a *Lambek network*. In other words, a Lambek network is an information network where the signaling relation coincides with the composition operation thought of as a three-place relation. Here are two natural examples of Lambek networks.

Example 7 Let $S = C$ consist of the finite strings on some alphabet Σ. Define $\circ$ to be concatenation of strings and consider the associated Lambek network. Thus the noun phrase *Mary* is the channel between the transtive verb *admires* and the verb phrase *admires Mary*. The latter is, in turn, the channel between the noun phrase *John* and the sentence *John admires Mary*.

Example 8 The connection with the Lambek Calculus can be further illuminated by considering the relational semantics for the Lambek Calculus suggested by van Bentham and shown to be complete in [Andréka and Mikulás 1994]. Here models are taken to be the ordered pairs in some transitive relation R. To construe such a relation as an information network, we take the sites and channels both to be the elements of R and define $\langle a, b \rangle \circ \langle b, c \rangle = \langle a, c \rangle$. We add a new element u (for "undefined") and define $\langle a, b \rangle \circ \langle c, d \rangle = u$ if $b \neq c$. We also define $x \circ u = u \circ x = u$ for all x. This makes $\circ$ an associative operation. If we consider the associated Lambek network we have $\langle a, b \rangle \overset{\langle b, c \rangle}{\rightsquigarrow} \langle a, c \rangle$ for all pairs $\langle a, b \rangle$ and $\langle b, c \rangle$ in the transitive relation R and for no other pairs. That is, we think of a pair $\langle b, c \rangle$ in R as a channel which takes as source any pair in R whose second element is b, say $\langle a, b \rangle$, and connects it to a unique target, namely $\langle a, c \rangle$.

As a final example, we consider action and plans. This example was suggested by work of Bibel and associates [Bibel et al. 1989] applying linear logic to planning.

Example 9 Suppose we have some first order language L_0 with relations and constants. For our sites, we take L_0-structures over some fixed domain A. For channels, we take arbitrary relations between such structures, i.e., sets of pairs $\langle \mathcal{M}_l, \mathcal{M}_r \rangle$ of such structures. We can think of these pairs $\langle \mathcal{M}_l, \mathcal{M}_r \rangle$ as models for a language L_1 which is just like L_0 except that every predicate R of L_0 has two versions R_l and R_r in L_1.

To be a little more concrete, the language L_0 might be used to describe blocks worlds, having predicates like $\mathsf{Block(x)}$, $\mathsf{Table(x)}$, $\mathsf{Empty(x)}$, and $\mathsf{On(x,y)}$. The channels we would be interested in are relations that capture actions like the action of moving a onto b, or simply moving a somewhere. Thus, for example, the channel $MoveOn(a, b)$ would consist of those pairs $\langle \mathcal{M}_l, \mathcal{M}_r \rangle$ such that $\mathcal{M}_l$ and $\mathcal{M}_r$ are just alike except that a is moved onto b in $\mathcal{M}_r$. (We assume that this can only succeed if b was empty in M_l. We also assume that the stack of blocks that were on a stay on it as it is moved onto b.) Thus, for example, the L_1-sentence

$$\forall x \forall y \, [On_r(x,y) \supset (On_l(x,y) \vee x = a)]$$

expresses a truth about the action of moving a. (We use $\supset$ for the material conditional since we want to save $\rightarrow$ for the conditional of information flow.)

We are primarily interested in *information flow* along chains of connections

$$s \overset{c_1}{\rightsquigarrow} t_1 \overset{c_2}{\rightsquigarrow} t_2 \rightsquigarrow \cdots \overset{c_n}{\rightsquigarrow} t_n$$

If we have such a chain and we know various things about elements of the chain, what can we tell about other elements of the chain? The simplest case of this is a chain of length 1, $s \overset{c}{\rightsquigarrow} t$. If we know that s is of such-and-such a type and that c is of such-and-such a type, what can we tell about t?

4.2 A Language of Types

In order to have a theory of information, we need to have ways of classifying sites, so that we can say that if site s is of type A and if it is connected to some site t by a channel c that supports the inference $A \rightarrow B$, then t is of type B. We can think here of these "types" syntactically, as expressions in some language, or more semantically, as units of information. But in either case we need a calculus of types and an analysis of what it means for a site or channel to be of some type.

Our language has some basic types and four connectives for building more complex types: $\downarrow, \rightarrow, \leftarrow$ and $\circ$. These are read as follows, where we use A, B to range over types of sites; C and D to range over types of channels.

$A \downarrow C$ is read "A and C"
$A \leftarrow C$ is read "A given C."
$A \rightarrow B$ is read "A to B."
$C \circ D$ is read "C and then D."

Both of these "and" connectives are noncommutative, but for different reasons, as we will see.

The historical neglect of channels in logic no doubt reflects an intuition that channels are, somehow, of a different nature than the things they connect, what we call sites (sources and targets) in this paper. And in some of our examples, sites and channels are disjoint from one another. But some of the other examples, most notably those illustrating the Lambek calclus, but also Example 6 are important cases where channels are themselves particulars with their own channels connecting them to other particulars. There are, then, two options to be explored, accordingly as we impose sorting constraints in the language. One option reflects the first kind of example, where sites and channels are kept separate. The other is more appropriate when channels are themselves sites. In the full paper (Barwise et al. 1994) on which this is a report, we explore both options. Here we discuss only the second of these two.

We introduce a single-sorted language L with a collection *AtExp* of atomic types. An *information model* $\mathcal{M}$ for L consists of an information network $\mathcal{N}$ together with a function f assigning to each $A \in \mathit{AtExp}$ some set of sites and channels. If $x \in f(A)$ we will say that x is of type A in $\mathcal{M}$. The expressions of L are given by the following context-free grammar:

$$\mathit{Exp} =: \mathit{AtExp} \mid (\mathit{Exp} \downarrow \mathit{Exp}) \mid (\mathit{Exp} \leftarrow \mathit{Exp}) \mid (\mathit{Exp} \rightarrow \mathit{Exp}) \mid (\mathit{Exp} \circ \mathit{Exp})$$

In words, this amounts to the following recursive definition of the set of expressions:

- Every atomic expression is an expression.
- If $A \in \mathit{Exp}$ and $B \in \mathit{Exp}$, then $(A \downarrow B)$, $(A \leftarrow B)$, $(A \rightarrow B)$ and $(A \circ B)$ are in *Exp*.

We call expressions *types* since we are thinking of them as classifying sites and channels.

The basic intuition is that $A \rightarrow B$ classifies those channels between sites that take one from a site of type A to a site of type B. A channel is of type $A \circ B$ if it can be decomposed into a channel of type A followed by one of type B. These intuitions are captured by the definition of satisfaction in a model given below. First, though, we present some examples.

Example 10 For a first example, consider the network where (Gödel numbers of) computable functions connect natural numbers to natural numbers. For site types we might take sets of natural numbers, like EVEN, PRIME, and ODD. If e is a Gödel number of the function $3x + 1$, then e will be of type EVEN→ODD, read "even to odd," since the function takes any even number x to an odd number. It will also be of type ODD→EVEN. However, it will not be of type PRIME→EVEN since it connects the prime number 2 to the number 7 which is not even. A number n will be of type PRIME ↓(ODD→EVEN) if n is of the form $f(m)$ where m is prime and f is a recursive function that takes odds to evens. (Every number is of this type, of course, since we can always send 2 to n and everything else to some even number.)

Example 11 For another example, consider the network of strings under concatenation given in Example [7]. For atomic types, we might take English syntactic categories like N, TV, VP and S. Then the English expression

Mary would be of type both N and of type TV$\to$VP (the type of expression that takes a TV to its left and gives back a VP).

Example 12 Consider the case of action and plans, as modeled in Example 9. Here it is natural to use the first-order sentences of L_0 as basic types over the sites and first-order sentences of L_1 as basic types over the channels (actions). Thus we would have basic types like $\mathsf{MoveOnto(a,b)}$ or $\mathsf{Move(a)}$ to classify actions. The new connectives will allow us to form sentences like $\mathsf{Empty(b)} \downarrow \mathsf{Move(a,b)}$, which would hold at a model M iff this model can be gotten from a model where b was empty by the action of moving a onto b. An action will be of type $\neg\mathsf{Empty(b)} \to \mathsf{Empty(b)}$ if it takes one from situations where b has something on it to situations where b is empty. An action will be of type

$$(\neg\mathsf{Empty(b)} \to \mathsf{Empty(b)}) \circ \mathsf{Move(a,b)}$$

if it consists of some action that uncovers the covered block b, followed by moving a onto b. Notice that this is quite different than

$$\mathsf{Move(a,b)} \circ (\neg\mathsf{Empty(b)} \to \mathsf{Empty(b)})$$

which holds of those actions which first move a onto b and then empty b. The first entails that the resulting situation is of type $\mathsf{On(a,b)}$ while the second entails just the opposite. An example of the $\leftarrow$ connective, consider $\mathsf{On(a,b)} \leftarrow \mathsf{Move(a)}$, read "On(a,b) given Move(a)". This will classify those situations where a will be on b, given any action of moving a. In order for this to hold, it must be that the table is full and b is the only block that is empty (except for a or the block that is on the top of the stack above a). Notice that whereas $A \to B$ combines sentences $A, B \in L_0$, $A \leftarrow B$ combines a sentence A of L_0 with a sentence B of L_1.

We now formalize the definition of what it means for a site or channel to be of some type. We use "s" and "t" to range over sources and targets and "c", possibly with subscripts, to range over channels.

Definition 2 Given an information model $\mathcal{M} = \langle \mathcal{N}, f \rangle$, the of-type relation $\models$ is defined by the following clauses:

- For an atomic type A, $x \models_{\mathcal{M}} A$ iff $x \in f(A)$
- **To:** $c \models_{\mathcal{M}} (A \to B)$ iff $\forall s, t$ (if $s \models_{\mathcal{M}} A$ and $s \overset{c}{\leadsto} t$, then $t \models_{\mathcal{M}} B$)
- **And then:** $c \models_{\mathcal{M}} (A \circ B)$ iff $\exists c_1, c_2$($c_1 \models_{\mathcal{M}} A$, $c_2 \models_{\mathcal{M}} B$ and $c = c_1 \circ c_2$)
- **And:** $t \models (A \downarrow C)$ iff $\exists s, c$ ($s \models_{\mathcal{M}} A$, $c \models_{\mathcal{M}} C$ and $s \overset{c}{\leadsto} t$)
- **Given:** $s \models_{\mathcal{M}} (A \leftarrow C)$ if and only if $\forall c, t$ (given that $c \models_{\mathcal{M}} C$, if $s \overset{c}{\leadsto} t$, then $t \models_{\mathcal{M}} A$)

Both $\downarrow$ and $\circ$ are forms of conjunction. There are three ways to see this. First, we can just look at the definition and see that "and" appears in both. Second, we note that if the information network is the trivial "one world" network of Example 2, then both of these amount to ordinary conjunction. Finally, the claim is further substantiated by the inference rules for these connectives in the calculi below. By adding certain structural rules, the first rules for $\downarrow$ and $\circ$ would degenerate into the rules for truth-functional "and." In a similarly manner, we can see that $\rightarrow$ and $\leftarrow$ are generalizations of the material conditional.

The following example is worth noting.

Example 13 Let $\mathcal{N}$ be the one-world network of Example 2, let $f(A) = \{w\}$ for every basic type A, and let $\mathcal{M} = \langle \mathcal{N}, f \rangle$ be the resulting model. An easy induction shows that $w \models_{\mathcal{M}} A$ for every type A. This is because our language does not have any form of negation.

With our calculus of types at hand, there is a decision to be made in giving an analysis of information flow. We can think of the types themselves as units of information, leaving the sites and channels that support them implicit. Alternatively, we can think of the information units as being given by a pair $[s : A]$ consisting of a site s (or channel) and a type A. In this case, the logic will traffic directly in such units. We explore both alternatives in the full version of this paper. In the present version, we discuss only the former.

4.3 Two Kinds of Sequents

We want to develop a Gentzen calculus for the language developed in the previous section. Thus we want to know what it means for a sequent $A_1, \ldots, A_n \vdash B$ to be valid. However, as it turns out, there are two reasonable notions, having to do with the two different functions a given object might assume: that of a site or that of a channel. These two functions give us two distinct notions of what it means for a sequent to be valid in a model.

Consider, for example, the sequents:

$$A, A \rightarrow B \vdash B$$

$$A \rightarrow B, B \rightarrow C \vdash A \rightarrow C$$

While these both look quite reasonable at first sight, a second's thought shows that they are only reasonable under different interpretations of $\vdash$. The first is valid if what $\vdash$ means is that if s is a site of type A and c is a channel that connects s to t and c is of type $A \rightarrow B$ then t is of type B. The second is valid if what it means is that if c and d are channels of type $A \rightarrow B$ and $B \rightarrow C$ respectively, then $c \circ d$ is of type $A \rightarrow C$.

This suggests that for each sequence Γ of types and each expression A, we distinguish *two* sequents, an "s-sequent" $\Gamma \vdash A$ and a "c-sequent" $\Gamma \mid\!\sim A$.

Definition 3

- An s-sequent $A, C_1, \ldots, C_n \vdash B$ is *valid in* a model $\mathcal{M}$ if and only if for every information chain

  $$s \overset{c_1}{\leadsto} t_1 \overset{c_2}{\leadsto} t_2 \leadsto \cdots \overset{c_n}{\leadsto} t_n,$$

  if $s \models A$ and $c_i \models C_i$ for each i, then $t_n \models B$.
- A c-sequent $C_1, \ldots, C_n \mid\!\sim C$ is *valid in* $\mathcal{M}$ if for every sequence $c_1, \ldots, c_n$ of channels, if $c_i \models C_i$ for each i, then $(c_1 \circ \ldots \circ c_n) \models C$.

With this definition of validity, the first and fourth of the following are valid, whereas the middle two will be invalid.

$$A, A \to B \vdash B$$

$$A, A \to B \mid\!\sim B$$

$$A \to B, B \to C \vdash A \to C$$

$$A \to B, B \to C \mid\!\sim A \to C$$

Besides axiomatizing the set of valid sequents, we also want to axiomatize the notion of logical consequence between sequents, which we now define in the natural way.

Definition 4 A *theory* T is a set of sequents. A sequent S is a logical consequence of a theory T if S is valid in every model $\mathcal{M}$ in which all the sequents in T are valid. We write this as $\models_T S$.

Notice that every theory in our language is consistent, in virtue of Example 13. Thus axiomatizing the notion of consequence cannot be reduced to the problem of consistency.

Example 14 Let's look at a couple of valid sequents in the Example 9, the one where channels are actions. In this case, an s-sequent $A, C_1, \ldots, C_n \vdash B$ holds in a model if whenever you start with a situation s of type A and carry out actions of types $C_1, ..., C_n$, in that order, then whatever situation t you get to will be of type B. For instance part of the theory of our blocks world would be the sequent

$$\mathsf{Empty(b), MoveOn(a, b)} \vdash \mathsf{On(a, b)}$$

If we are given site types A and B and asked to devise a plan for getting from situations of type A to those of type B, what we want is an action type C such that $A, C \vdash B$ holds.[1]

[1]This is a bit crude. A better definition would be to find a constraint $A' \to B'$ such

A c-sequent $C_1, \ldots, C_n \mathrel{|\!\!\sim} C$ asserts that if you compose any actions of types $C_1, ..., C_n$, in that order, then the resulting action will be of type C. Thus, for example, in our blocks world model, we would have the

$$\mathsf{MoveOn(a, b), MoveOn(a, c)} \mathrel{|\!\!\sim} \mathsf{MoveOn(a, c)}$$

Notice, however, that if we permute the two premises, the result is not valid.

The problem of refining a plan C is the problem of finding ways to bring about an action of type C by composing actions of other types, $C_1, ..., C_n$, types that you can implement more directly. For instance, in going home, you leave your office, walk to your car, and drive home. Each of these is similarly refined until you get types of actions that you can actually carry out. Thus, the task of refing an action of type C can be modeled as the task of finding a valid c-sequent with C as succedent.

Example 15 Recall that a Lambek network is one where the signaling relation is the same as the composition operation thought of as a three-pace relation. The following are valid in every Lambek network, for all expressions A and B:

$$A \downarrow B \vdash A \circ B$$

$$A \circ B \vdash A \downarrow B$$

We call the set of all such sequents the *Lambek theory.* In a theory which includes the Lambek theory, the distinction between $\downarrow$ and $\circ$ is lost.

4.4 A Gentzen System

We are now ready to present the Gentzen system for our language. The system is a refinement of the Lambek calculus.

---

that $A \vdash A'$ and $A, A' \to B' \vdash B$ are both valid. The first sequent insures that the initial situation s of type A is guaranteed to satisfy the preconditions of the action to be undertaken.

(Identity) $A \vdash A$

(Application) $(\downarrow L)$ $\dfrac{A, B, \Gamma \vdash C}{A \downarrow B, \Gamma \vdash C}$ $(\downarrow R)$ $\dfrac{\Gamma \vdash A \quad \Delta \mid\!\sim C}{\Gamma, \Delta \vdash A \downarrow C}$

(Right Impl.) $(\rightarrow L)$ $\dfrac{\Gamma \vdash A \quad B, \Delta \vdash C}{\Gamma, A \rightarrow B, \Delta \vdash C}$ $(\rightarrow R)$ $\dfrac{A, \Gamma \vdash C}{\Gamma \mid\!\sim A \rightarrow C}$

(Left Impl.) $(\leftarrow L)$ $\dfrac{\Gamma \mid\!\sim C \quad A, \Delta \vdash B}{A \leftarrow C, \Gamma, \Delta \vdash B}$ $(\leftarrow R)$ $\dfrac{\Gamma, A \vdash B}{\Gamma \vdash B \leftarrow A}$

(Composition) $c-(\circ L)$ $\dfrac{\Gamma, A, B, \Delta \mid\!\sim C}{\Gamma, A \circ B, \Delta \mid\!\sim C}$ $(\circ R)$ $\dfrac{\Gamma \mid\!\sim A \quad \Delta \mid\!\sim B}{\Gamma, \Delta \mid\!\sim A \circ B}$

$s-(\circ L)$ $\dfrac{\Gamma, A, B, \Delta \vdash C}{\Gamma, A \circ B, \Delta \vdash C}$ $(\Gamma \neq \emptyset)$

Notice that when the sequence $\Gamma = \langle A \rangle$ is of length one, then the sequents $\Gamma \vdash B$ and $\Gamma \mid\!\sim B$ are semantically equivalent in that one holds in a model if and only if the other does, since they both say that everything of type A is also of type B. These are not derivable from the rules presented so far. Thus, to the above rules, we also need to have a rule that tells us that in the case where the sequence on the left of the sequent is a single type, these two notions coincide. That is, we need rules

(Trivial) $\dfrac{A \vdash B}{A \mid\!\sim B}$ $\dfrac{A \mid\!\sim B}{A \vdash B}$

We also include Cut as a basic rule, in three forms.

$$\frac{\Gamma \vdash A \quad A, \Delta \vdash B}{\Gamma, \Delta \vdash B} \qquad \frac{\Gamma \mid\!\sim A \quad \Delta, A, \Pi \vdash B}{\Delta, \Gamma, \Pi \vdash B} \qquad \frac{\Gamma \mid\!\sim A \quad \Delta, A, \Pi \mid\!\sim B}{\Delta, \Gamma, \Pi \mid\!\sim B}$$

Soundness and Cut Elimination It is routine to check that the above rules are sound, in the sense that if the premises of a rule are valid in a model, so is the conclusion. Hence, any sequent that is provable from a theory is a logical consquence of that theory. In the case of the empty theory, the Cut rule is not needed for the proof of completness given below, so the rule of Cut could, in this case, be eliminated. One can also prove a cut-elimination result directly.

4.5 Completeness

In this section we sketch the completeness of our system.

Definition 5 A model $\mathcal{M}$ of a theory T is a *characteristic model* of T if every sequent that is valid in $\mathcal{M}$ is provable from T.

As an immediate consequence of the soundness of our system, we note that if $\mathcal{M}$ is a characteristic model of T then the sequents which are valid in $\mathcal{M}$ are exactly those sequents which are provable from T. The completeness of our Gentzen system is an immediate consequence of the following result:

Theorem 1 *Every theory has a characteristic model.*

Corollary 2 (Completeness) *A sequent S is a logical consequence of a theory T iff it is provable from T in the above system.*

An earlier version of our results was weaker in that we had to resort to networks where composition was multiple-valued. The recent proof of the Completeness Theorem for the Lambek Calculus, relative to the relational semantics due to Andréka and Mikulás 1994 inspired the proof of this stronger result.[2] The proof also shows the following.

Theorem 3 *Every extension of the Lambek theory has a characteristic model whose network is a Lambek network.*

To sketch a proof of Theorem 1, let us fix a set T. We want to show how to construct a characteristic model for T.

The model $\mathcal{M}$ will be of the form $\langle \mathcal{N}, f \rangle$, where the network $\mathcal{N}$ is constructed as the limit of a sequence of "partial networks" $\mathcal{N}_n$ for $n < \omega$. At each stage we will throw in at most one new site (or channel) s (or c) and declare such an s to be labeled by some expression A. There will, in general, be many sites labeled by a given expression, not just one. Our aim is to make sure that for any type B, a site labeled by A is of type B if and only if $A \vdash B$ is provable from T. Thus, for example, if we label some site s by $A \downarrow C$ then we will make sure to throw in a site s_0 labeled by A and a connnection c labeled by C, and declare $s_0 \overset{c}{\rightsquigarrow} s$.

The principal obstacle to carrying out this construction involves composition. If we have labelled a channel expression c by some c-expression C and it happens that $C \vdash_T A \circ B$ is provable, then at some stage we need to throw in new channels c_0, c_1, label them by types A and B respectively, and define $c_0 \circ c_1 = c$. The trick is to do this in a way that insures that the final composition operation is associative.

Definition 6 A partial binary function $\circ_0$ on a set A_0 is said to be a *partially associative operation* on A_0 if there is a set A with $A_0 \subseteq A$ and an associative operation $\circ$ on A such that $\circ$ is an extension of $\circ_0$.

In order to prove the completeness theorem, we build up an information network in stages. At each stage, we need to make sure that the partial function which approximates our final composition operation, has not im-

[2] Jerry Seligman has constructed a clever alternate proof of our result, one which derives it from the completeness of the system in Andréka and Mikulás 1994 This proof is not yet written down however

plicitly forced us to identify channels which are distinct. That is, we need to make sure that the operation is a partially associative operation.

There is a standard construction in algebra that tells us when a partial binary function is partially associative. Given a partial operation $\circ$ on a set A, one takes the free semigroup on A (finite sequences over A under concatenation), and factors out by the smallest equivalence relations that identifies strings up to association, and that identifies strings forced to be identical by $\circ$. Providing no elements of A are identified in this factorization, then this construction gives one an associative extension of $\circ$.

While this construction is quite standard, we need to go into a bit more detail, in order to define a notion we need in the main lemma. Let A be a set with a partial binary operation $\circ$ defined on A. We want to characterize when $\circ$ is a partially associative operation on A. We write (xy) for the ordered pair $\langle x, y\rangle$. Let A^* be the smallest set containing A and closed under ordered pairs. We use α, β, γ to range over A^*.

We define four relations on A^*.

- α *expands to* β if there are elements $a, b, c \in A$ such that $a \circ b = c$ and β can be obtained from α by replacing one occurrence of c by the ordered pair (ab).
- α *contracts to* β if there are elements $a, b, c \in A$ such that $a \circ b = c$ and β can be obtained from α by replacing one occurrence of (ab) by the ordered pair c.
- α *regroups to* β if there are elements $\gamma_1, \gamma_2, \gamma_3 \in A^*$ such that β can be obtained by replacing one occurrence of one of the following by the other: $((\gamma_1\gamma_2)\gamma_3)$, $(\gamma_1(\gamma_2\gamma_3))$.
- Finally, we say that α *rewrites to* β, written $\alpha \asymp \beta$, iff $\alpha = \beta$ or there is a finite sequence $\alpha_1, \ldots, \alpha_n$, $1 \leq n$, such that $\alpha = \alpha_1$, $\beta = \alpha_n$, and for each $i < n$, α_{i+1} can be obtained from α_i by expansion, contraction, or regrouping. Such a sequence is called a rewrite sequence.

Notice that $\asymp$ is symmetric on A^* since expansion and contraction on converses of one another and regrouping is symmetric. It is also reflexive and transitive, and hence an equivalence relation on A^*.

Proposition 4 *(Extension Lemma) For any structure $\langle A, \circ\rangle$, where $\circ$ is a partial binary operation on A, the following are equivalent:*

1. *For all $a, b \in A$, if $a \asymp b$ then $a = b$.*
2. *$\circ$ is a partially associative operation on A.*
3. *There is an "initial" associative structure $\langle A', \circ'\rangle$ extending $\langle A, \circ\rangle$. That is, $\circ'$ is an extension of $\circ$, it is total on A', associative, and*

for any other such $\langle A'', \circ'' \rangle$, *there is a unique homomorphism* f *from* $\langle A', \circ' \rangle$ *into* $\langle A'', \circ'' \rangle$.

The following is obvious but quite important in what follows.

Lemma 5 $\circ$ *is a partially associative operation on* A *if and only if for all* $a, b, c \in A$, *if* $(ab) \asymp c$ *and* $a \circ b$ *is defined then* $a \circ b = c$.

While the condition given in this lemma seems a bit more complicated than condition (1) of the Extension Lemma, it is actually more useful for our purposes. The reason is that it allows us to make the following definition, and is why we needed to review this construction in the first place.

Definition 7 Let $\circ$ be a partial function on A and let $\circ^*$ be a subfunction of $\circ$. We say that $\circ^*$ is an *expansion basis* for $\circ$ if for all $a, b, c \in A$, if $(ab) \asymp c$ then there is a rewrite sequence from (ab) to c where the expansion rule *expand* x *to* (yz) is used only if $z = x \circ^* y$.

That is, in the rewriting, we need only expand z to some (xy) if the smaller function $\circ^*$ warrants the expansion.

The construction We are now ready to outline the construction of our information network $\mathcal{N}$ and then our model $\mathcal{M} = \langle \mathcal{N}, f \rangle$. Let S be a, countable infinite set. We will draw our sites and channels from S. Whenever we add a site s (or channel c) to our model, we will label it by an expression $A_s = \ell(s)$ (or $A_c = \ell(c)$) of the appropriate sort with the intent described above. The network $\mathcal{N}$ will be the union of an increasing chain of structures $\mathcal{N}_n = \langle Si_n, Ch_n, \rightsquigarrow_n, \circ_n \rangle$, for $n < \omega$. The structures will not themselves be information networks since the operation $\circ_n$ will be partial. But the limit will be an information network. We list various conditions that we will want to satisfy in building this sequence of structures. We identify each condition by an ordered tuple containing the key parameters in the condition.

Code	Condition
$\langle s, A, B \downarrow C\rangle$	If $A \vdash B \downarrow C$ is provable from T and A is the label of s then there is an s_0 labeled by B and a c labeled by C such that $s_0 \overset{c}{\rightsquigarrow} s$.
$\langle s, c, A \downarrow B\rangle$	If A is the label of s and B is the label of c then there is a $t \in S$ labeled by $A \downarrow B$ and $s \overset{c}{\rightsquigarrow} t$.
$\langle c, A, B \circ C\rangle$	If $A \vdash B \circ C$ is provable from T and A is the label of c then there is a c_0 labeled by B and a c_1 labeled by C such that $c = c_0 \circ c_1$.
$\langle c_0, c_1, B \circ C\rangle$	If B is the label of c_0 and C is the label of c_1 then there is a $c \in Ch$ such that $c_0 \circ c_1 = c$ and $A_c \vdash B \circ C$ is provable from T.
$\langle s, t, c, d, e\rangle$	If s, t are sites, c, d, e are channels, $e = c \circ d$, and $s \overset{e}{\rightsquigarrow} t$ then there is a site r such that $s \overset{c}{\rightsquigarrow} r$, $r \overset{d}{\rightsquigarrow} t$, and $A_r \vdash_T A_s \downarrow A_c$ and $A_t \vdash_T A_r \downarrow A_d$.
$\langle s, r, t, c, d, e\rangle$	If s, r, t are sites, c, d, e are channels, $e = c \circ d$, and $s \overset{c}{\rightsquigarrow} r$ and $r \overset{d}{\rightsquigarrow} t$ then $s \overset{e}{\rightsquigarrow} t$.

Using standard techniques from cardinal arithmetic, order these tuples in a list of order type ω, say $\tau_0, \tau_1, \ldots, \tau_n, \ldots$, so that each tuple occurs infinitely often. We will examine the condition associated with τ_n at stage n of our construction.

Lemma 6 *(Main Lemma) There is an increasing sequence of structures $\mathcal{N}_n = \langle Si_n, Ch_n, \rightsquigarrow_n, \circ_n, \circ_n^*, \ell_n\rangle$, for $n < \omega$, satisfying the following conditions:*

1. *The composition operation $\circ_n$ is a partially associative operation on Ch_n, with an expansion basis $\circ_n^*$.*
2. *ℓ_n is a function from $Si_n \cup Ch_n$ into expressions. If $s \in Si_n$ then $\ell_n(s)$ is an expression of sort* **s**, *called the label of s, and written as A_s. Similarly, if $c \in Ch_n$ then $\ell_n(c)$ is an expression of sort* **c**, *called the label of c, and written as A_c.*
3. *The condition associated with the tuple τ_i is satisfied in $\mathcal{N}_n$ for all $i < n$.*
4. *If $s \overset{c}{\rightsquigarrow} t$ in $\mathcal{N}_n$ then $A_t \vdash_T A_s \downarrow A_c$*
5. *If $c = c_0 \circ_n c_1$ then $A_c \vdash_T A_{c_0} \circ A_{c_1}$*
6. *If $c = c_0 \circ_n^* c_1$ then $A_{c_0} \circ A_{c_1} \equiv_T A_c$.*

Lemma 7 *Given any sequence as in Lemma 6, let $\mathcal{N} = \bigcup_{n<\omega} \mathcal{N}_n$. Then structure $\mathcal{N}$ is an information network. That is, $\circ$ is a total, associative operation and $s \overset{c \circ d}{\rightsquigarrow} t$ if and only if there is a site r such that $s \overset{c}{\rightsquigarrow} r$ and $r \overset{d}{\rightsquigarrow} t$.*

To turn our information network $\mathcal{N}$ into a model $\mathcal{M} = \langle \mathcal{N}, f \rangle$, define, for each atomic expression B,

$$f(B) = \{s \in S \mid A_s \vdash_T B\}$$

where A_s is the label of the site s.

Lemma 8 *In the model $\mathcal{M}$ just constructed, $s \models B$ iff $A_s \vdash_T B$, and similarly for channels.*

Finally, we need to show that each sequent in the theory T is valid in $\mathcal{M}$ and that an unprovable sequent is not valid in $\mathcal{M}$. To begin, let's assume that $S \in T$. We assume S is an s-sequent, the other case being similar. We may suppose that S is of the form $A \vdash B$. Let s be any site in $\mathcal{M}$ such that $s \models A$. Then by the lemma, $A_s \vdash A$. But then by cut, $A_s \vdash B$ and so, again by the lemma, $s \models B$. Now let us show that if S is valid in $\mathcal{M}$, then $A \vdash_T B$. Let s be any site labeled by A. Hence, if S is valid in $\mathcal{M}$, then $s \models B$. But then $A \vdash B$ by Lemma 8.

4.6 Conclusions

From the point of view of information flow, the Lambek calculus is rather restrictive in that it considers only Lambek networks, that is, networks where sites and channels are the same thing and the only connections between two, say a and c, is determined by whether there is a b such that $a \circ b = c$. Moving from Lambek network to arbitrary networks makes for a framework more suitable to the general study of information flow.

However, once one makes this move, it become clear that there are a lot of additional connectives one could profitably study, and other approaches one could follow to pursue the logic of information flow. We pursue some of these in the fuller version of this paper. Many others remain unexplored.

References

Andréka, Hajnal, and Szabolcs Mikulás. 1994. Lambek Calculus and its Relational Semantics: Completeness and Incompleteness. *J Logic, Language, and Information* 3:1–37.

Barwise, J., and J. Seligman. 1993. Imperfect Information Flow. In *Proceedings of the 8th Annual IEEE Symposium on Logic in Computer Science*, ed. Moshe Vardi. 252–261. Los Alamito, CA.: IEEE Computer Society Press.

Barwise, Jon. 1993. Constraints, Channels, and the Flow of Information. In *Situation Theory and Its Applications, III*, ed. Peter Aczel, David Israel, Yosuhiro Katagiri, and Stanley Peters. 3–27. CSLI Lecture Notes Number 22. Stanford: CSLI Publications.

Barwise, Jon, Dov Gabbay, and Chrysafis Hartonas. 1994. On the Logic of Information Flow. Technical report. Indiana University Logic Group.

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes.* Bradford Books. Cambridge, Mass.: MIT Press.

Bibel, W., L. Fariñas del Cerro, B. Fronhöfer, and A. Herzig. 1989. Plan Generation by Linear Proofs : On Semantics. In *GWAI'89, 13th German Workshop on Artificial Intelligence*, ed. D. Metzing, Informatik–Fachberichte, Vol. 216, 49–62. Schloß Eringerfeld, Geseke, Germany, September. Springer, Berlin.

Lambek, J. 1958. The Mathematics of Sentence Structure. *The American Mathematical Monthly* 65:154–170.

5

Logical Aspects of Combined Structures

PATRICK BLACKBURN AND MAARTEN DE RIJKE

Abstract

This is an exploratory paper about combining structures. Typically when one applies logic to such areas as computer science, artificial intelligence or linguistics, one encounters hybrid ontologies. The aim of this paper is to identify plausible strategies for coping with such ontological richness.

Introduction

This is an exploratory paper about combining structures. The need for various such combinations has come up in many areas, including computer science (Aceto 1992, Montanari et al. 1993), artificial intelligence (Hobbs 1985), linguistics (Blackburn et al. 1993, 1994a), philosophy (Seligman and Barwise 1993) and logic itself (Kracht and Wolter 1991, de Rijke 1993).

The aim of this paper is to identify the issue of combining structures (and of combining logics and theories, for that matter) as a new research line. We do this as follows. We first present a list of examples in Section 1. In Section 2 we introduce a very simple framework for combining structures using so-called *trios*; briefly, a trio is a triple consisting of a two classes of structures and a collection of links between them. We give examples of theories of specific trios, and we discuss how properties of structures that are combined into trios, transfer — or don't transfer — to the trio. Section 3 concludes the paper with a discussion of further questions.

A final introductory remark: this paper is a preliminary report of ongoing work; a fuller account will be given in (Blackburn and de Rijke 1994).

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

5.1 Examples

In this section we present examples. We focus on combining structures, but much of what we will say below can be couched in terms of combining logics or theories. We start with two simple examples of what we call *refinement semantics* in which one ontology is given additional structure at the atomic level by other ontologies; we then move on to the richer *classification semantics* in which one structure classifies the elements of another structure by inducing an equivalence relation on it. Finally we consider *fully interacting structures*, where there is no restriction on the relation between the structures being combined.

Generalized Phrase Structure Grammar

Our first example of a refinement semantics stems from generative linguistics: the Generalized Phrase Structure Grammar of Gazdar *et al.* (1985) views linguistic structure as a combined ontology, namely finite trees fibered over finite feature structures, that is: finite trees such that to every node in the tree is associated a finite labeled transition system in which every transition relation $\xrightarrow{a}$ is a partial function.

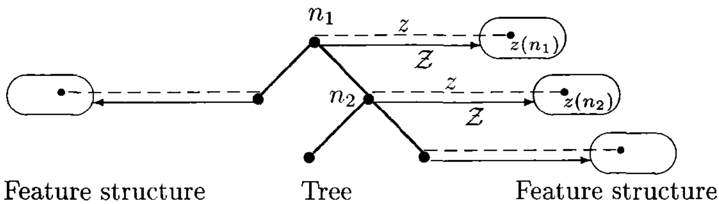


FIGURE 1 Linguistic structure in GPSG

In GPSG the feature structures are used to refine the notion of grammatical category. In contrast to the usual practice in formal language theory where the nodes of parse trees are decorated with 'indivisible' information about categories (for example NP for Noun Phrase or VP for Verb Phrase), GPSG splits the atom: an NP is now a structured object, a feature structure, that contains information about various subatomic features and values.

Finite trees fibered over finite feature structures provide a semantics for two distinct languages: a tree language $\mathcal{L}^T$ that moves us around the tree, and a feature language $\mathcal{L}^F$ that allows us access to the inner structure of grammatical categories. The central ideas of GPSG can then be expressed in a mixture of the two languages called $\mathcal{L}^T(\mathcal{L}^F)$ — the language $\mathcal{L}^T$ *layered over* the language $\mathcal{L}^F$ — in which the $\mathcal{L}^F$ wffs are viewed as the atomic wffs of $\mathcal{L}^T$. A wff ϕ in the layered language $\mathcal{L}^T(\mathcal{L}^F)$ is evaluated as follows: in general ϕ contains $\mathcal{L}^T$ connectives that move us around the tree until we hit what used to be the atomic level; instead of invoking an assignment or valuation at this stage we have further work to do: we *zoom in* from the tree node n to the associated feature structure $\mathcal{Z}(n)$ and start evaluating at $z(n)$.

Refinement is a very simple way of combining structures; the interaction between the components is limited — nodes in the feature structure, for example, simply don't have permission to access the tree structure. This restriction has a number of pleasant consequences; it's usually fairly straightforward to combine completeness and decidability results for the component logics into completeness and decidability results for the layered language (cf. Section 3 below).

Action Refinement in Process Theory

The previous example involving GPSG concerned refinement of states. In the present example we consider refinement of *actions* or *transitions*. In the top-down design of distributed systems one uses actions and states on an abstract level to represent complex processes on a more concrete level, leading naturally to refinement of states (as in the earlier GPSG example), and of actions (Aceto 1992).

Consider the design of an input device, repeatedly reading data and sending it off. A first, and highly abstract description is given in Figure 2.

FIGURE 2 An input device

On a slightly less abstract level of description the action 'read data' decomposes into 'prepare reading' and 'carry out reading.' This corresponds to Figure 3:

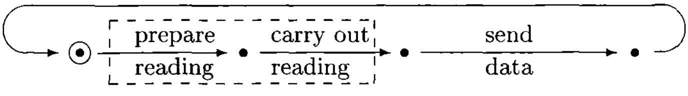


FIGURE 3 The input device refined

This is a very simple kind of refinement of actions: it just refines by a sequence of actions. In general more sophisticated types of refinement may be needed; one can think of refinement by parallel actions, or by infinite processes. This is best formulated as a form of substitution of structures in the following manner. Let r be a function from the (atomic) actions of a labeled transition system $\mathfrak{T}$ to rooted transition systems. The *refinement of* $\mathfrak{T}$ *by* r is the structure that is obtained as follows. For $s \xrightarrow{a} t$ an edge in $\mathfrak{T}$ let $r(a)'$ be a new copy of $r(a)$; identify s with the root of $r(a)$, identify t with all end nodes of $r(a)$, and remove the edge $s \xrightarrow{a} t$. In other words, instead of making an a-transition at s, we now start at the root of $r(a)'$, traverse a terminating path through $r(a)'$, and then continue from t onwards.

In passing, it's quite natural to look systematically at the converse of refinement: *abstraction*. One could take a structure $\mathfrak{A}$ to be an abstraction of a structure $\mathfrak{B}$ if $\mathfrak{A}$ is the quotient of $\mathfrak{B}$ under an appropriate notion of morphism (see Hobbs (1985) for an example use of abstraction in AI). A general approach would allow for refinements/abstractions over any kind of item in ones structures simultaneously.

Lexical Functional Grammar

In many applications where structures or logics need to be combined, more complex interactions are required than refinements; *classifications* provide an important example of such combinations. To explain these, and to see how classifications are different from refinements, it's best to return to generative grammar; more specifically, we will look at Lexical Functional Grammar (LFG) (Kaplan and Bresnan 1982). Like GPSG, LFG views syntactic structure in terms of composite entities made from finite trees and finite feature structures, but it glues these together rather differently. The basic picture is the one given in Figure 4.

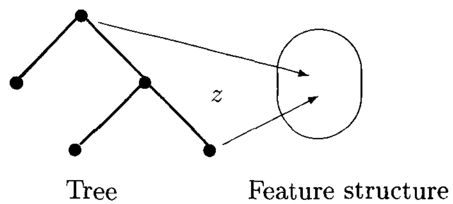


FIGURE 4 Linguistic structure in LFG

Here we have a single finite tree and a single finite feature structure linked by a partial function z. This feature structure induces a classification of tree nodes via z in the following sense. According to LFG, sentences embody two levels of structure: constituent structure, which is represented by a tree, and grammatical relations, represented by a feature structure. Then, two tree nodes are identified, or classified as 'being functionally the same,' if they are mapped onto the same point in the feature structure.

Note that this is *not* the same as refinement of atomic information, rather it's about ensuring that the internal structures of the two ontologies correctly 'match' each other. LFG enforces the required matching using phrase structure rules annotated with equations. For example, $\uparrow = \downarrow$ means that if we move up the tree from a node t, and then zoom in to the feature structure, we arrive at the same point we would have reached by zooming in directly from t.

Channel Theory

Another area where the idea of using one structure to classify the objects of another is Situation Semantics. Situation semantics has long emphasized the importance of ontological diversity, and one branch where this is put forward very elegantly is the *channel theory* initiated by Jerry Seligman (1990).

As part of a general attempt to model laws or regularities, and information flow, a classification is defined as a triple $\mathbf{A} = \langle tok(\mathbf{A}), typ(\mathbf{A}), : \rangle$, where $tok(\mathbf{A})$ and $typ(\mathbf{A})$ are non-empty sets (of tokens and types, respectively), and : is the *classification relationship* between tokens and types (see Figure 5).

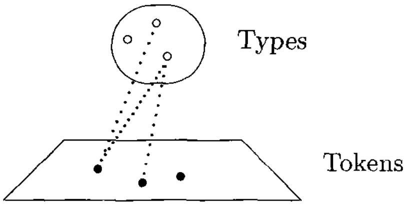


FIGURE 5 A classification

Here the types of $\mathbf{A}$ classify the tokens of $\mathbf{A}$, and the types induce a natural equivalence relation $\sim$ of indistinguishability on tokens: $a \sim b$ iff for all types α we have $a : \alpha$ iff $b : \alpha$. As with LFG and its annotated phrase structure rules, further restrictions may be imposed on the way types and tokens interact.

In channel theory classifications are not considered in isolation. A further 'stacking' of structures occurs when classifications are combined into so-called channels to model information flow. A channel is something which directs information flow between classifications. This is achieved as follows. First, a notion of information preserving morphisms between classifications $\mathbf{A}$ and $\mathbf{B}$ is defined as a certain kind of bi-function $f : \mathbf{A} \rightrightarrows \mathbf{B}$. Then, a channel $\mathbf{C} : \mathbf{A} \Longrightarrow \mathbf{B}$ is a classification $\mathbf{C}$ together with morphisms $\text{left}_C : \mathbf{C} \rightrightarrows \mathbf{A}$ and $\text{right}_C : \mathbf{C} \rightrightarrows \mathbf{B}$.

Roughly, the tokens of $\mathbf{C}$ are used to model connections between the tokens of $\mathbf{A}$ and the tokens of $\mathbf{B}$, and the types of $\mathbf{C}$ are used to express constraints between the types of $\mathbf{A}$ and the types of $\mathbf{B}$; and a connection is

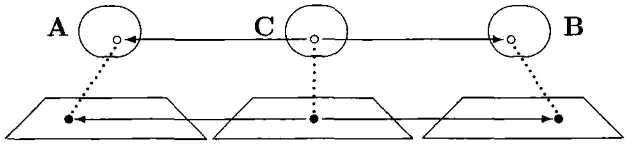


FIGURE 6 A channel

classified by a constraint just in case information flows along the connection in a way that conforms to the constraint.

Full Interaction: Fibering

In an essential way the four examples of combining structures or logics given so far all involve only *one way* traffic between structures: objects in one structure convey information about objects in another structure. In a number of recent talks and papers Dov Gabbay has advocated the idea of *fibering* two sets of semantic entities over each other (Gabbay 1991). Roughly, a fibered structure consists of two classes of models, each class with its own language, plus a function between the classes that tells you how to evaluate formulas belonging to the one language inside structures of the other.

To make this more concrete, here is an example: we fiber finite trees and finite equivalence relations; for the sake of this example we assume that we have two mono-modal languages, $\mathcal{L}_T$ for talking about trees, and $\mathcal{L}_E$ for talking about equivalence relations.

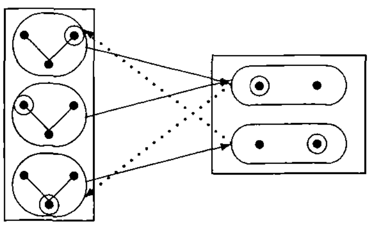


FIGURE 7 Fibering a tree and an equivalence relation

First of all, let a *model-state* pair be a pair $(\mathfrak{M}, s)$ where $\mathfrak{M}$ is a model based on a finite tree or on a finite equivalence relation, and s is an element of $\mathfrak{M}$. Second, let M_T, M_E be non-empty sets of model-state pairs whose first component is a finite tree or a finite equivalence relation, respectively, and such that if $(\mathfrak{M}, s) \in \mathsf{M}_T \cup \mathsf{M}_E$ and $s' \in \mathfrak{M}$, then $(\mathfrak{M}, s') \in \mathsf{M}_T \cup \mathsf{M}_E$. Now, for the fibering function, let F be a pair of functions (F_T, F_E) with $F_T : \mathsf{M}_T \to \mathsf{M}_E$ and $F_E : \mathsf{M}_E \to \mathsf{M}_T$ such that model-state pairs that are mapped onto each other agree on all atomic symbols common to both languages; the fibering function regulates the interaction between the classes of structures M_T and M_E. Finally, for F a fibering function, the F-fibered structure over M_T and M_E is the triple (W_F, R_F, V_F) such that

- W_F is $\mathsf{M}_T \cup \mathsf{M}_E$,
- R_F is $\{\, ((\mathfrak{M}_1, s_1), (\mathfrak{M}_2, s_2)) : \mathfrak{M}_1 = \mathfrak{M}_2 \text{ and } Rs_1s_2 \,\}$,
- V_F is simply the union of the component valuations.

As to the evaluation of complex formulas, tree formulas are interpreted in M_T as usual, and likewise for $\mathcal{L}_E$-formulas and M_E. If we hit a tree formula while evaluating in M_E, we apply the fibering function F to the current model-state pair, and continue evaluating in its associated model-state pair in M_T; a similar move is made when we hit an $\mathcal{L}_E$ subformula while evaluating in M_T.

We should point out that more involved definitions of fibering (or similar constructions) have been proposed in the literature (Eiben et al. 1992,Gabbay 1991, Goguen and Burstall 1984). For our purposes the definition given here suffices.

Much contemporary research in logic is strongly influenced by applications — and not merely the traditional applications in philosophy or mathematics. Instead, new interdisciplinary work in such areas as Cognitive Science, Artificial Intelligence and Theoretical and Computational Linguistics is the focus of attention. This broadening of the scope of applied logic forces the logician to take ontological diversity seriously, and emphasizes the need for investigations such as the present one.

5.2 Trios

In this section we present a first pass at a mathematical framework for combining structures.

Definition 1 Let A and B be two classes of structures, and let Z be a collection of relations between the elements of A and those of B. Then the triple $(\mathsf{A}, \mathsf{Z}, \mathsf{B})$ is called a *trio*. The classes A and B are called the *left* and *right continents*, respectively, of the trio, and Z is called its *bridge*.

As an example, the trios in GPSG style refinement consist of a single tree $\mathfrak{A}$ as their left continent, a right continent consisting of $|\mathfrak{A}|$ many structures $\{\, \mathfrak{B}_a : a \text{ in } \mathfrak{A} \,\}$, and a bridge consisting of an injective function linking each point of $\mathfrak{A}$ to an element of the right continent.

Of course, the general notion of a trio will only lead to useful and interesting theorizing when we refine it. Such refinements can be pursued along at least two lines. First of all we can try to develop the systems theory of trios.

How do they combine? What kind of structure do they form? It seems that 2-categories (Street 1987) will provide a natural setting for understanding trios at a very abstract level. This kind of question is left to a separate paper.

Here we pursue a second line of questions that comes up in connection with trios: logical issues. We will touch upon logics of specific trios, the analysis of specific bridges, and the classification of bridges.

Logics of specific trios. Consider a *bisimilar* trio $(\mathfrak{A}, \underline{\leftrightarrow}, \mathfrak{B})$ where $\mathfrak{A}$, $\mathfrak{B}$ are labeled transition systems with transition relations $\xrightarrow{a}$ and $\xrightarrow{b}$, respectively, and $\underline{\leftrightarrow}$ is a bisimulation between $\mathfrak{A}$ and $\mathfrak{B}$; that is, $\underline{\leftrightarrow}$ is a non-empty relation on $\mathfrak{A} \times \mathfrak{B}$ that only relates points with the same atomic information and that satisfies a back-and-forth condition: if $x, y \in \mathfrak{A}$, $x' \in \mathfrak{B}$, $x \xrightarrow{a} y$ and $x \underline{\leftrightarrow} x'$, then there is a $y' \in \mathfrak{B}$ such that $x' \xrightarrow{b} b'$ and $y \underline{\leftrightarrow} y'$ (and likewise in the opposite direction). We also assume that $\mathfrak{A}$ comes with a mono-modal language $\mathcal{L}(\langle a \rangle)$, and $\mathfrak{B}$ with a language $\mathcal{L}(\langle b \rangle)$.

A first decision we have to make is: what language do we use to talk about such bisimilar trios? Given that we have components $\mathfrak{A}$, $\mathfrak{B}$ and $\underline{\leftrightarrow}$, the natural set-up has two constants `left` and `right` to denote $\mathfrak{A}$ and $\mathfrak{B}$ respectively, and four diamonds $\langle a \rangle$, $\langle b \rangle$ and $\langle z \rangle$, $\langle z^{-1} \rangle$, where $\langle z \rangle$, $\langle z^{-1} \rangle$ let us move back-and-forth between the two continents $\mathfrak{A}$ and $\mathfrak{B}$, that is: they are interpreted using the bisimulation relation $\underline{\leftrightarrow}$.

An obvious question is: what is the logic of bisimilar trios? — We need at least the axioms and rules of inference of the minimal modal logic **K** for each of $\langle a \rangle$, $\langle b \rangle$, $\langle z \rangle$ and $\langle z^{-1} \rangle$. In addition the following axioms should be added:

- the well-known axioms from temporal logic stating that the interpretation of $\langle z \rangle$ is the converse of the interpretation of $\langle z^{-1} \rangle$;
- $\texttt{left} \vee \texttt{right}$ and $\neg(\texttt{left} \wedge \texttt{right})$ to force every point to live in exactly one continent;
- $\phi \leftrightarrow (\phi \wedge \texttt{left})$ for all $\mathcal{L}(\langle a \rangle)$ formulas, and likewise with `right` and $\mathcal{L}(\langle b \rangle)$ formulas, to force the interpretation of `left` and `right` to be an $\mathcal{L}(\langle a \rangle)$ model and an $\mathcal{L}(\langle b \rangle)$ model, respectively;
- $\langle z \rangle\phi \to \texttt{left} \wedge \langle z \rangle(\texttt{right} \wedge \phi)$ and $\langle z^{-1} \rangle\phi \to \texttt{right} \wedge \langle z^{-1} \rangle(\texttt{left} \wedge \phi)$ to force the interpretation of $\langle z \rangle$ to be a subset of 'the interpretation of `left` $\times$ the interpretation of `right`';
- $\texttt{left} \wedge p \to [z]p$ and $\texttt{right} \wedge p \to [z^{-1}]p$ (*for p atomic!*), to force the condition on atomic information;
- $\langle a \rangle\phi \to [z]\langle b \rangle\langle z^{-1} \rangle\phi$ and $\langle b \rangle\phi \to [z^{-1}]\langle a \rangle\langle z \rangle\phi$, to force the back-and-forth conditions.

Theorem 1 *The above set of axioms and rules completely axiomatizes validity of bisimilar trios.*

The proof of the theorem is a canonical model construction; as the requirement that $\leftrightarrow$ be non-empty is not expressible we may have to tinker somewhat with the canonical model — but this can be done using standard techniques from modal logic.

Analyzing specific bridges. We now move on to a slightly more general *genre* of question. Fix a kind of bridge, and let Z be a bridge of that kind — what do we know about completeness, decidability, complexity ... of the trios $(\mathsf{A}, \mathsf{Z}, \mathsf{B})$, given that we have complete, decidable ... theories for the continents? Here are a few examples.

Finger and Gabbay (1992) prove some general transfer results for a special form of the notion of refinement that we discussed in Section 2. They show how to add a temporal dimension to a logic system, or in our terms: they take temporal logic with *Since* and *Until* over the natural numbers as the 'top language', and refine the atomic information of that language, using any other language as the 'bottom language'. The results Finger and Gabbay establish include that, provided the bottom language has a complete axiomatization, the combined language has one as well; and, provided the bottom language is decidable, so is the combined one. In the full paper we enhance and generalize these transfer results in a number of ways. First, we also consider transfer and non-transfer of complexity results. Second, we show that the Finger and Gabbay results remain valid when other (one-dimensional) top languages are used instead of *Since*, *Until* logic. And third, to capture phenomena such as Action Refinement in Process Theory (as discussed in Section 2), we consider transfer problems for top languages whose formulas are interpreted at semantic objects other than single states, including pairs, transitions, and sequences.

As a second example, following their introduction in the formal semantics of natural language, Shehtman (1978) considers the *Cartesian product* of two modal logics. For instance, the intended frames of the Cartesian product of the modal logics **S4** and **S5** consists of structures whose universe is a product $U_0 \times U_1$ with a pre-order on U_0 and an equivalence relation on U_1. An important question here is to determine in which cases $\mathbf{L}(\mathfrak{F}_1 \times \mathfrak{F}_2) = \mathbf{L}(\mathfrak{F}_1) \times \mathbf{L}(\mathfrak{F}_2)$, that is, when does the logic of the product coincide with the product of the component logics? Shehtman (1978) provides a partial answer. Another important example of a similar 'simple' combination of structures arises when we consider so-called independent joins of logics. For instance, the independent join of two mono-modal logics $\mathbf{L}_1$ and $\mathbf{L}_2$ with distinct modal operators $\langle a \rangle$ and $\langle b \rangle$, respectively, is simply the union of the two logics. On the level of structures this operation amounts to considering structures $(W, \xrightarrow{a}, \xrightarrow{b})$ that have reducts living in the language of $\mathbf{L}_1$ and in the language of $\mathbf{L}_2$. Kracht and Wolter (1991) show that the independent join of two complete or decidable logics is again complete or decidable. And Spaan (1993) completely classifies the complexity of the independent join of two modal logics in terms of the complexity of the component logics; she also analyzes the much more difficult situation in which the component logics are not fully independent. In the full paper we present further results along these lines.

Classifying bridges. The final of type of question we want to mention is still more general; it can roughly be summarized as: what kind of bridges are there? Questions like this serve a dual purpose: on the one hand they spring from a desire to find sensible ways of cutting up the universe of all trios (in Section 2 we only considered three kinds of trios: refinements, classifications and full interactions); on the other hand having available a taxonomy of trios and bridges may help us in obtaining generalizations — and thus in gaining a better understanding — of the results obtained so far. For example, this may help us to understand why refinements and independent joins behave nicely. These issues are the focus of our ongoing technical investigations.

5.3 Discussion

At both the technical and conceptual level there is much obvious work to do. For example, the completeness result for bisimilar trios is just a pointer to further results; Blackburn and de Rijke (1994) axiomatize other logics of specific trios, and indeed it is possible to state and prove general completeness results for trios in the spirit of Sahlqvist's Theorem.

It seems hard to state general results and properties of combined structures without moving to a very abstract mathematical framework. As has already been mentioned, to understand the systems theory of trios at a general level, we feel that 2-categories may be useful. However, for particular kinds of trios dedicated system theories can be much more appropriate; channel theory as a theory of clasification structures provides an example.

To conclude the paper let us consider a very obvious weakness of the story we have told so far; we have acted as if combined ontologies are lifeless, *static* entities. This ignores the fact that for many applications it is precisely the *dynamic* aspects of combined ontologies that are of interest. To make matters more concrete, we revert to generative grammar. Consider Tree Adjoining Grammars (tags) (Joshi et al. 1975). Tag analyses are essentially dynamic; sentences are viewed as the result of merging trees together. To gain something of the flavour of tags in action, consider the operation known as *adjunction*. Let τ be a tree with an internal node labeled by the nonterminal symbol A. Let ρ be an auxiliary tree with root and foot node labeled by the same nonterminal symbol A. The tree τ' that results by adjoining ρ at the A-labeled node in τ is formed by removing the subtree of τ rooted at this node, inserting ρ in its place, and substituting it at the foot node of ρ. Perhaps the most important thing to notice is the role played by the node labeled A. We began with an initial structure (namely τ) with a designated node (namely that labeled A); we then performed a computation step; and this created a larger structure with a new designated node, the site for further creation. Of course, all this *could* be described statically. But to do so does violence to the underlying intuitions. We need analyses which cope with the growth of structures rather than merely treating them as completed objects.[1] This idea brings us to territory already explored by much of the literature on feature logic (Carpenter 1992), on evolving algebras (Gurevich 1991) and on specification languages (Groenboom and Renardel de Lavalette 1994). Ultimately this seems to require investigations of 'imperative logics', that is, logics that write to models rather than treating them as read-only structures; see Blackburn, de Rijke and Seligman (1994b) for some preliminary investigations.

Acknowledgments

Both authors would like to thank the Netherlands Organization for Scientific Research (NWO), project NF 102/62-356 'Structural and Semantic Parallels between Natural Languages and Programming Languages' for financial support.

References

Aceto, L. 1992. *Action Refinement in Process Algebras.* Cambridge: Cambridge University Press.

Blackburn, P., and M. de Rijke. 1994. Zooming In, Zooming Out. Technical report. CWI, Amsterdam. To appear in *Journal of Logic, Language and Information.*

Blackburn, P., M. de Rijke, and C. Gardent. 1994a. Back and Forth through Time and Events. In *Temporal Logic (ICTL '94)*, ed. D.M. Gabbay and H.J. Ohlbach. 225–237. LNAI 827. Berlin: Springer.

Blackburn, P., M. de Rijke, and J. Seligman. 1994b. Modal Logics with Memory. Unpublished manuscript.

Blackburn, P., C. Gardent, and W. Meyer Viol. 1993. Talking about Trees. In *Proceedings 6th EACL.* 21–29. Utrecht.

Carpenter, B. 1992. *The Logic of Typed Feature Structures.* Cambridge: Cambridge University Press.

de Rijke, M. 1993. What is Modal Logic? In *Arrow Logics and Multi-Modal Logics*, ed. L. Pólós, M. Masuch, and M. Marx. Oxford: Oxford University Press. To appear.

[1]Although this is precisely what domain theory sets out to do, it remains to be seen whether domain theoretic tools are the most suitable setting for the applications considered in this paper.

Eiben, A., A. Jánossy, and Kurucz. Á. 1992. Combining Logics. Technical Report IR-319. Amsterdam: Department of Mathematics and Computer Science, Free University.

Finger, M., and D.M. Gabbay. 1992. Adding a Temporal Dimension to a Logic System. *Journal of Logic, Language and Information* 1:203–233.

Gabbay, D.M. 1991. Labelled Deductive Systems, 6th intermediate draft. Lecture Notes 3rd ESSLLI Summer School, Saarbrücken.

Gazdar, G., E. Klein, G. Pullum, and I. Sag. 1985. *Generalized Phrase Structure Grammar.* Oxford: Basil Blackwell.

Goguen, J.A., and R.M. Burstall. 1984. Some Fundamental Algebraic Tools for the Semantics of Computation (I). *Theoretical Computer Science* 31:175–209.

Groenboom, R., and G. Renardel de Lavalette. 1994. Reasoning about Dynamic Features in Specification Languages. Unpublished manuscript.

Gurevich, Y. 1991. Evolving algebras; a tutorial introduction. *Bulletin EATCS* 43:264–284.

Hobbs, J. 1985. Granularity. In *Proceedings 9th IJCAI. Vol. 1*, 432–435.

Joshi, A., L. Levy, and M. Takahashi. 1975. Tree Adjunct Grammars. *Journal of Computer and Systems Sciences* 10.

Kaplan, R., and J. Bresnan. 1982. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In *The Mental Representation of Grammatical Relations*, ed. J. Bresnan. 173–281. Cambridge, MA: MIT Press.

Kracht, M., and F. Wolter. 1991. Properties of Independently Axiomatizable Bimodal Logics. *Journal of Symbolic Logic* 56:1469–1485.

Montanari, A., E. Ciapessoni, E. Corsetti, and P. San Pietro. 1993. Dealing with Time Granularity in Logical Specifications of Real-Time Systems. Technical report. Department of Mathematics and Computer Science, University of Udine.

Seligman, J. 1990. *Perspectives: a Relativistic Approach to the Theory of Information.* Doctoral dissertation, University of Edinburgh.

Seligman, J., and J. Barwise. 1993. Channel Theory: Toward a mathematics of imperfect information flow. Unpublished manuscript.

Shehtman, V. 1978. Two-Dimensional Modal Logic. *Matematicheskie Zametki* 23:759–772. (In Russian).

Spaan, E. 1993. *Complexity of Modal Logics.* Doctoral dissertation, ILLC, University of Amsterdam.

Street, R. 1987. The algebra of oriented simplexes. *Journal of Pure and Applied Algebra* 49:283–335.

6

On Rich Ontologies for Tense and Aspect

PATRICK BLACKBURN, CLAIRE GARDENT AND MAARTEN DE RIJKE

Abstract

In this paper *back-and-forth structures* are defined and applied to the semantics of natural language. Back-and-forth structures consist of an event structure and an interval structure communicating via a relational link; transitions in the one structure correspond to transitions in the other. Such entities enable us to view temporal constructions (such as tense, aspect, and temporal connectives) as methods of moving systematically between information sources. We illustrate this with a treatment of the English present perfect, and progressive aspect, that draws on ideas developed in Moens and Steedman (1988), and discuss the role of rich ontologies in formal semantics.

Introduction

Formal accounts of temporal constructions in natural language often disagree about the semantic ontology to be assumed — should it be point based, interval based or event based? We think that more adequate analyses of natural language will be obtained by *combining* ontologies, not choosing between them. We illustrate this by combining interval structures with (various forms of) event structures into what we call *back-and-forth structures* (BAFs). These consist of an interval structure and an event structure linked by a relation so that transitions in the one correspond to transitions in the other.

Such combined ontologies enable us to build our analyses round the

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

following intuition: temporal constructions are means of systematically exploiting links between information sources. Consider the English present perfect. It is common to informally gloss this construction as 'a past tense of present relevance'. For example, '*John has gone to the store*' means that at some past time John went to the store and, moreover, that John's excursion is somehow of relevance to the present context. We see two important transitions here: a move backwards in time through an interval structure, and a move to an associated event in an event structure. The English present perfect coordinates these transitions, and BAFs enable us to model this.

Much of this abstract uses BAFs to explore the ideas of Moens and Steedman (1988); indeed, BAFs developed by thinking about the kind of machinery required to formalise their work. Moens and Steedman provide a wide ranging account of temporal semantics (topics considered include tense, temporal reference, aspect and adverbial modification) couched as a Winograd-style procedural semantics. Their work hinges on (at least) the following ideas: that non-temporal relations between events must be admitted if an adequate account is to be given of the semantics of '*when*' and various aspectual phenomena; that there are key event configurations (called 'nuclei') underlying the richness of event ontology; and that adverbial (and other forms of) modification are to be accounted for in terms of 'type coercion'. The Moens and Steedman account is attractive because while it is wide ranging, its explanations reduce to the interaction of a handful of intuitive ideas. Its weakness is that it is largely unformalised. We believe BAFs provide a setting in which substantial parts of their account can be made precise. BAFs can be seen as a way of modeling the insight that a systematic interplay between temporal and non-temporal relations is called for, and by progressively enriching the event structures they are built over one can model ever more of the Moens and Steedman system.

We proceed as follows. We informally discuss the semantics of the English present perfect, indicating why the use of combined ontologies seems promising. We then introduce *simple BAFs*. These consist of interval structures combined with an extremely simple type of eventuality structure. Although such structures are too simple to cope with all the subtleties of natural language, their use permits the central idea underlying our proposal to be clearly presented. Following this, we (slightly) enrich the eventuality component to form *sorted BAFs*. This enables us to refine our discussion of the present perfect, and to provide an analysis of progressive aspect that does not run foul of the so-called imperfective paradox. To conclude the paper we describe how we are extending this work, discuss some methodological issues (why should one be interested in this style of semantic analysis?) and note some other BAF-like proposals we have found in the literature.

6.1 The Present Perfect

While descriptive work on the English present perfect abounds, the construction has been notoriously resistant to formal analysis. In this section we discuss the problems the present perfect gives rise to, and argue that these indicate the need for combined ontologies.

It is often argued that the English present perfect is used to describe past events of present relevance. Perhaps the most well-known account of this intuition is that described in Reichenbach (1947), where a present perfect is analysed as describing a past event (the event temporally precedes the speech time) whose reference time coincides with the speech time. Reichenbach's reference point is meant to be the temporal perspective from which the described event is viewed. Because reference and speech time coincide in the the present perfect, this tense enables one to present a past event as being of relevance to the present. This contrasts with the simple past which is viewed as describing a past event whose reference time coincides with the event time rather than with the speech time.

Although Reichenbach's approach goes one step toward capturing the intuition underlying the use of the present perfect, two problems remain. First, what is the nature of reference times, and how are they determined? Second, the Reichenbachian account fails to account for many observations made in the literature concerning the restrictions governing the use of the present perfect. For instance, it does not explain why the sentence in (1) is infelicitous if uttered at a time occurring after the coffee has been cleaned.

(1) I have spilled my coffee.

Similarly, it does not account for the restrictions placed by verbal aspectual classes on the use of the present perfect, for example:

(2) a. ? The house has been empty (stative expression)
b. ? I have worked in the garden (process expression)
c. ? The star has twinkled (point expression).

Example (2a) shows that the present perfect is awkward in combination with stative expressions; (2b) and (2c) illustrate its awkwardness in combination with process expressions and point expressions, respectively.

As Moens and Steedman (1986) convincingly argue, these problems can be resolved if the internal structure of events is taken into account. Briefly, the idea is that an event (or *nucleus* in Moens and Steedman's terminology) is a tripartite structure consisting of a *preparatory phase*, a *culmination* and a *consequent state*. Given such a structure, the function of the present perfect is to situate the reference time in the consequent state of the core event being described (cf. Moens and Steedman (1986), p.20). Thus instead of the Reichenbach schema

```
   E              R, S
---|--------------|-----
```

Moens and Steedman describe the present perfect by means of the following diagram:

```
                E         R, S
|---------------|----------------
       PP      Cul         CS
```

Their account incorporates the central Reichenbachian intuition, while eliminating its problematic aspects:

- The reference point is given a (more) precise and more motivated location in time, namely within the time stretch of the consequent state.
- Example (1) is explained as follows. An obvious consequence of spilling one's coffee is that coffee is spilled. Under the Moens and Steedman theory, uttering a sentence in the present perfect indicates (i) that the reference time coincides with the speech time and (ii) that both these times are included in the time stretch of the consequent state. Thus by uttering the present perfect (1), the speaker indicates that coffee is still spilled. Hence the oddity of (1) in a context where it isn't.
- The ill-formedness of the examples in (2) is explained by the fact that stative, process and point expressions are used to describe either states (i.e. unstructured entities) or these parts of the event structure which do not include the consequent state.[1] Since these expressions do not involve the notion of consequent state, they cannot be used in the present perfect whose semantics is defined in terms of this very notion.

The Moens and Steedman approach is intuitively appealing: how can it be made precise? We believe this can be done quite straightforwardly by combining ontologies.

Intuitively, their approach demands a mixture of ontologies: at the very least it seems to call for *temporal* structure, *eventuality* structure, and (crucially) a 'sensible fit' between these two ontologies. The 'past tense' component of the present perfect seems to require some notion of temporal structure; at the very least, this will involve some notion of temporal precedence. But this temporal structure does not suffice: in addition we need to

[1] These aspectual notions are discussed in more detail in Section 4.

invoke some notion of 'eventuality', and some sort of relation of 'relevance' between eventualities (for example, between the act of spilling the coffee, and the presence of the coffee on the floor). Intuitively this relevance relation isn't temporal; nonetheless, capturing the idea that we want an event of *present* relevance seems to presuppose that some sort of 'synchronisation' between the precedence relation on the temporal structure and the relevance relation on the eventuality structure is in force.

Actually, we will need even more structure than this. As examples (2a)–(2c) showed, the present perfect does not willingly combine with all verb types. We will need to work with a suitably fine-grained view of eventuality structure to capture these restrictions; in particular, by using eventuality structures *sorted* in a manner that reflects verbal aspectual classes we can model more of the Moens and Steedman account.

In the following sections we will present two simple formal models that capture some of these intuitions. We first present *simple* BAFs. These combine interval structures with a very simple notion of eventuality structure in a way that permits the intuition of 'present relevance' to be directly captured. (Or, to put it in the terminology of Moens and Steedman, they enable us to model the intuition that the present perfect works by locating the reference point in the run-time of consequent state induced by the eventuality being described.) We then refine this simple picture by enriching the eventuality structures used to make BAFs. The resulting *sorted* BAFs allow us to model the aspectual restrictions governing the use of the present perfect, and yields a simple solution to the imperfective paradox.

6.2 Simple BAFs

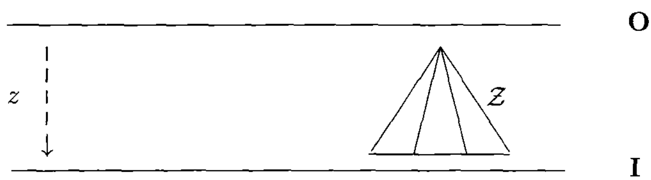
Simple BAFs consist of four components: an *interval structure*, an *eventuality structure*, and (most importantly) two *links* between them.

An *interval structure* **I** is a triple $\langle I, <, \sqsubseteq \rangle$ as defined in van Benthem (1991). Here I is a set of intervals, $<$ is the precedence relation, and $\sqsubseteq$ is the subinterval relation. We work with *linear*, *atomic* interval structures. That is, we assume that given any two intervals either one precedes the other or they overlap, and that our structures contain minimal, 'point-like' intervals.

An *eventuality* structure of signature $\mathcal{E}$ is (for the purposes of the present section) a triple $\mathbf{O} = \langle O, \texttt{GRiTo}, \{P_e\}_{e \in \mathcal{E}} \rangle$. Here O is a non-empty set, the set of *eventuality occurrences*; GRiTo is a binary relation on O; and all the P_e are unary relations on O. We assume $\mathcal{E} \neq \emptyset$. If e GRiTo e' then we say e *gives rise to* e'. The unary relations P_e can be thought of as 'eventualities' for example *runnings*, *jumpings* and *recitings of poems*.

Now the crucial step. A *back-and-forth structure* (BAF) of signature $\mathcal{E}$

is a quadruple $\langle \mathbf{O}, z, \mathcal{Z}, \mathbf{I} \rangle$, where $\mathbf{O}$ is an eventuality structure of signature $\mathcal{E}$, $\mathbf{I}$ is an interval structure, z is a function from O to I that returns the run-time or temporal extent of an eventuality and that preserves the relation $\mathtt{GRiTo}$: if e $\mathtt{GRiTo}$ e' then $z(e) < z(e')$. That is, z is an order-preserving morphism from the eventuality structure to the interval structure; it is this morphism that synchronizes the two ontologies. $\mathcal{Z}$ is the relation with domain O and range I defined by $e\mathcal{Z}i$ iff $i \sqsubseteq z(e)$. That is, we assume that all eventualities are *downward persistent* to subintervals.



We now formulate a toy language for talking about BAFs: its vocabulary consists of all the items in $\mathcal{E}$, which we shall write as p, q, r, ... etc., and call *eventuality symbols*, and an operator PERF. If α is an eventuality symbol then PERF α is well formed (and nothing else is). Obviously it would be possible to add the Boolean operators and allow arbitrary embeddings of PERF; but while this leads to fairly interesting logical territory, it has little relevance to the semantics of natural language.

Now for the semantics. Let $\mathbf{B}$ $(= \langle \mathbf{O}, z, \mathcal{Z}, \mathbf{I} \rangle)$ be a BAF. Then, for all intervals i, and all eventuality symbols q, we define:

$$\begin{array}{lll} \mathbf{B}, i \models \textsc{Perf}\, q & \text{iff} & \exists i' \exists e' \exists e \big(i' < i \ \& \\ & & i' = z(e') \ \& \\ & & e' \in P_q \ \& \\ & & e' \ \mathtt{GRiTo}\ e \ \& \\ & & e\mathcal{Z}i \big). \end{array}$$

Consider what this does. Suppose we have a sentence in the present perfect, say '*Fire has broken out on the oil rig*'. In our toy language this takes the form:

PERF(*Fire breaks out on the oil rig*).

If we evaluate this at an interval i in $\mathbf{B}$, then we must 'complete a square' in a BAF back to the utterance interval i. That is, we move back in time to an interval i' which is the run-time for an event e'; this e' is an eventuality of the correct type (that is, e' is a breaking out of a fire) and moreover e' gives rise to an event e which is $\mathcal{Z}$ related to our utterance interval i. Intuitively, the eventuality of present relevance e would be the ongoing burning of the fire, that is the consequent state of the breaking out of the fire event. Roughly, this semantics relates to Reichenbach and

Moens and Steedman's approaches as follows: i is the time of speech (S), i' is the event time (E) and e is the consequent state induced by the event being described, namely e'. The Reichenbachian constraint according to which speech and reference times coincide is replaced by the Moens and Steedman intuition that the time stretch of the consequent state includes the speech time. In this way, we capture the intuition of present relevance which characterises the English present perfect.

6.3 Sorted BAFs

Simple BAFs have the virtue of making clear the fundamental idea underlying our approach, but they are very crude. To encode the aspectual restrictions placed on the use of the present perfect, and to model further temporal constructions such as the progressive, we need to say more about the relation between time and aspect. This is the object of the present section. We will insist that the eventuality structures used to make BAFs embody the sortal distinctions, and additional relations, demanded by the various verb classes. We start by motivating these additions.

Eventualities

On the basis of the tenses, aspects and adverbials with which they occur, we classify eventualities into five types; our classification is similar to the one of Carlson (1981) and Moens and Steedman (1988). First we distinguish between indefinitely extending eventualities which we call *states*, and eventualities with defined beginnings and ends called *events*. Sentence (3) describes a state:

(3) Her hair is black.

Events are subdivided into *atomic* and *extended* events, depending on whether or not their runtimes are an atomic interval.

To motivate a further subdivision of the extended events, compare sentences (4) and (5) below.

(4) Bert was writing a thesis.

(5) Bert was sleeping.

The difference between sentences such as (4) and sentences such as (5) has been observed by numerous authors, and is often couched in terms of accomplishments and activities, cf. Vendler (1967). We express this distinction between (4) and (5) by saying that the event reported in (4) has a natural *culmination*, viz. the completion of the thesis; (5) has no such culmination. Processes that tend to have culminations in this sense are said to be *culminating*. Both the accomplishments of Vendler (1967) and the culminat*ed* processes of Moens and Steedman (1988) are composite events, consisting of a culminating process and a culmination; we feel it is more natural to split those composites and refer explicitly to the completion relation between culminating processes and their culminations.

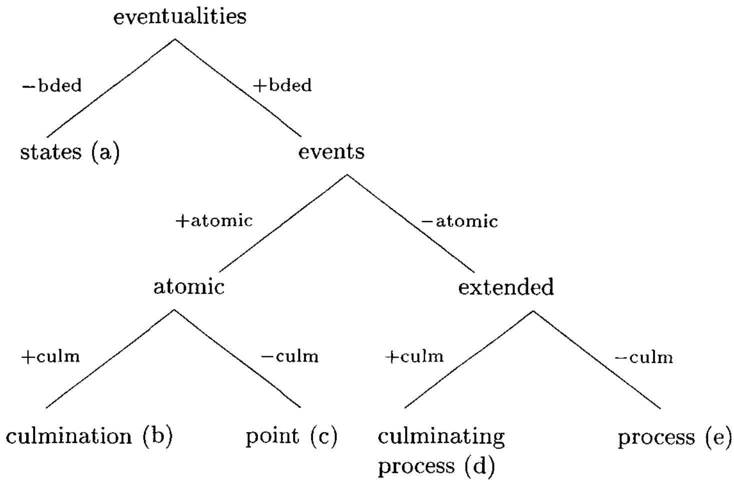
Corresponding to the above distinction between processes and culminating processes, we divide atomic events into points and culminations. They differ in that culminations describe the culmination of a structured event (or nucleus) whereas points simply describe isolated atomic events; as a result a culmination may be associated with a culminating process and a consequent state whereas points cannot. To understand this division consider sentences (6) and (7) below.

(6) Bert completed his thesis.

(7) Bert hiccupped.

Sentence (6) reports a culmination; its culminating process is the writing of the thesis, its consequent state a state where the thesis is completed. Without further 'world knowledge' no natural culminating process or consequent state can be associated with the point event of (7).

Here, then, is a scheme of the eventualities we distinguish:



Typical examples are:

(a) be green, know
(b) recognize, complete a paper
(c) hiccup, twinkle
(d) build a house, write a thesis
(e) play the piano, sleep, waste time

To sum up: the aspectual category of a sentence determines the sort of eventuality being described. Process, state and point expressions refer to some unstructured entity whereby a stative expression describes some unstructured event stretching over an unbounded period of time, a process expression some unstructured event stretching over a bounded period of time and a point expression some unstructured atomic event. In contrast, culminating process and culmination expressions are the building blocks of more structured eventualities (the 'nuclei') demanded by Moens and Steedman; these consist of a culminating process, a culmination and a consequent state appropriately linked. We now formally define our sorted structures and introduce the two sortally sensitive relations needed for building nuclei.

Sorting Eventuality Structures

We now extend the structures used earlier to incorporate these ideas. First, a *sorted eventuality structure* is a tuple

$$\mathbf{O} = \langle \mathtt{Point}, \mathtt{Culm}, \mathtt{Proc}, \mathtt{Culm\_Proc}, \mathtt{State}; \mathtt{GRiTo}, \mathtt{Compl}, \mathtt{Cons}; \{P_e\}_{e \in \mathcal{E}} \rangle,$$

where Point, Culm, Proc, Culm_Proc and State are mutually disjoint domains whose elements are used to interpret the various aspectual categories described above. GRiTo is just the 'gives-rise-to' relation defined in Section

3. We will continue to treat GRiTo as 'sortally insensitive'; that is, we will impose no restrictions on the sorts of the eventualities that can be included in its domain and range. The two new relations, Compl and Cons, are more interesting. Essentially, they are the first step in formalising the tripartite structures that underly the work in Moens and Steedman (1988). Triples (e, e', e'') such that e Compl e' and e' Cons e'' are Moens and Steedman style nuclei: e can be thought of as the preparatory process, e' as the culmination, and e'' as the consequent state. Let us examine these new relations more closely.

Compl is a binary relation (the *completion relation*) between culminating processes and culminations: it links a culminating process with its culmination. This motivates three further constraints on Compl. First, and most importantly, Compl must be a partial function: each culminating process can have at most one culmination. (As not all events which have a natural culmination actually reach it, we only have a *partial* function here. This will later yield a solution of the 'imperfective paradox'.) Secondly, we assume that for every culmination there is a culminating process that is linked by Compl to this culmination. (This simply demands that every culmination is the culmination of *something*; there are no stray culminations.) Thirdly, we assume that Compl is a subset of GRiTo. (Intuitively, if a culminating process has a culmination, it certainly gave-rise-to that culmination.)

Cons is a binary relation (the *consequences relation*) linking culminations with states. Intuitively, Cons links a culmination with the consequent state it gives rise to. This intuition motivates two further constraints on Cons: it should be a function, and it should be a subset of GRiTo. These restrictions seem to formalise the intentions underlying Moens and Steedman treatment of the link between culminations and consequent states. Roughly speaking, although a culmination might give-rise-to several consequent states, one of these is the 'preferred' or 'default' consequent state. The role of Cons is to 'select' this preferred consequence from the GRiTo relation. Moreover, every culmination gives rise to at least one consequent state (trivially, every 'winning of the race' gives raise to a state of 'having won the race') thus Cons is a *total* function.

To summarize:

1. GRiTo is a binary relation on $\mathtt{Point} \cup \mathtt{Culm} \cup \mathtt{Proc} \cup \mathtt{Culm\_Proc} \cup \mathtt{State}$
2. $\mathtt{Compl} \subseteq \mathtt{GRiTo}$.
3. Compl is a partial function whose domain is a subset of Culm_Proc and whose range is Culm.
4. $\forall e\,(\mathtt{Culm}(e) \rightarrow \exists e'\,(\mathtt{Culm\_Proc}(e')\ \&\ e'\,\mathtt{Compl}\,e))$.
5. $\mathtt{Cons} \subseteq \mathtt{GRiTo}$.

6. `Cons` is a total function whose domain is `Culm` and whose range is a subset of `State`.

Now that we now what sorted eventuality structures are, we can make richer BAFs. A *sorted BAF* is a BAF $\mathbf{B} = \langle \mathbf{O}, z, \mathcal{Z}, \mathbf{I} \rangle$, where $\mathbf{O}$ is a sorted eventuality structure in which the following additional conditions are satisfied:

4. $\forall e\,(\mathtt{Point}(e) \rightarrow z(e)$ is an atomic interval).
   $\forall e\,(\mathtt{Culm}(e) \rightarrow z(e)$ is an atomic interval).
5. $\forall e\,(\mathtt{Proc}(e) \rightarrow z(e)$ is an non atomic bounded interval).
   $\forall e\,(\mathtt{Culm\_Proc}(e) \rightarrow z(e)$ is an non atomic bounded interval).
6. $\forall e\,(\mathtt{State}(e) \rightarrow z(e)$ is an non atomic, non bounded interval).
7. $\forall e, i\,(i \sqsubseteq z(e) \leftrightarrow e\mathcal{Z}i)$.

Item 4 says that points and culminations are atomic events, item 5 that processes and culminating processes are non atomic bounded eventualities and item 6 that states are non atomic, unbounded eventualities; the seventh item ensures that eventualities are *downward persistent*. Note that BAFs *do* distinguish between points and culminations; only culminations can enter into the `Cons` relation. Similarly, the `Compl` relation differentiates between processes and culminating processes.

The Present Perfect and Sentence Aspect

In Section 2 we proposed a simple 'complete the square' semantics for the present perfect. Essentially, we used simple BAFs to formalise the intuition that the event talked about gives-rise-to some other eventuality (the consequent state) whose run time includes the speech time. We also observed that not all verbs may be naturally used with the present perfect. In this section we will see how sorted BAFs allow us to capture these distinctions. We *won't* be changing our semantics in any way; rather, we will just use the new (sortally sensitive) `Cons` relation to refine it.

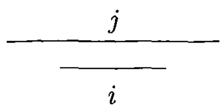
Consider sentences that are 'awkward' in the present perfect, such as *I have spilled my coffee* (a process sentence) or *The star has twinkled* (a point sentence). The key fact about such examples is that when uttered without any supporting context, there simply is no natural consequent state that can be associated with them. Conversely, given enough supporting context (say, a pair of coffee stained trousers, or a rhapsody on the stillness of an autumnal night) both sentences become acceptable. In short, neither process sentences nor point sentences 'inherently supports' the present perfect; but the construction can be used (and with exactly the semantics we discussed earlier) given suitable contextual support.

On the other hand, culminations 'inherently support' the present perfect, and sorted BAFs make it clear why. Consider an utterance of the culmination sentence *John has won the race.* Now — entirely irrespective of whether or not there is supporting context — John's winning of the race gives rise to at least one consequent state. This follows from the semantics provided by sorted BAFs. Let e be the event of John winning the race. This is an eventuality of sort culmination; that is, $e \in \texttt{Culm}$. But $\texttt{Cons}$ is a total function with domain $\texttt{Culm}$ and range $\texttt{State}$. Moreover $\texttt{Cons} \subseteq \texttt{GRiTo}$. Thus e gives-rise-to at least one consequent state, namely $\texttt{Cons}(e)$.

In short, whenever we hear a culmination sentence we expect a consequent state — and the sorted BAF semantics always provides a consequent state for culmination sentences via the $\texttt{Cons}$ function. Because they have a consequent state 'built in', culmination sentences are 'privileged users' of the present perfect construction.

Progressive Aspect and the Imperfective Paradox

We will now examine progressive aspect using sorted BAFs. Following Kamp and Reyle (1993), we assume that the function of the English progressive is to focus attention on the (culminating) process of some eventuality. This idea can be captured as follows. First, we enrich our toy language by adding the operators PAST and PROG, and allowing expressions of the form $\textsc{Past}\,q$ and $\textsc{Prog}\,q$ and $\textsc{Past}\,\textsc{Prog}\,q$ to be well formed. As for the semantics, first, define $i \sqsubseteq^+ j$ to hold between two intervals i, j if the following is the case:



Let $\mathbf{B}$ $(= \langle \mathbf{O}, z, \mathcal{Z}, \mathbf{I}\rangle)$ be a sorted BAF. Then, for all intervals i, we define the relation $\mathbf{B}, i \models \phi$ as follows:

$$\begin{array}{lll}
\mathbf{B}, i \models \textsc{Prog}\,q & \text{iff} & \exists e\,(e \in P_q \;\&\; (\texttt{Proc}(e) \vee \texttt{Culm_Proc}(e)) \\
 & & \qquad \&\; i \sqsubseteq^+ z(e)) \\
\mathbf{B}, i \models \textsc{Past}\,q & \text{iff} & \exists j, e\,(j < i \;\&\; e\mathcal{Z}j \;\&\; e \in P_q) \\
\mathbf{B}, i \models \textsc{PastProg}\,q & \text{iff} & \exists j\,(j < i \;\&\; \mathbf{B}, j \models \textsc{Prog}\,q).
\end{array}$$

One of the merits of such a semantics for the progressive is that it yields a simple solution to the so-called 'imperfective-paradox'. Following Dowty (1979), this paradox has been discussed by numerous authors. Briefly, the paradox is this: how can we account for the meaning of a progressive sentence like (8) and (10) in such a way that (8) may be true without (9) ever becoming true, while on the other hand (10) would tautologically imply (11)?

(8) Bert was writing a thesis.

(9) Bert wrote a thesis.

(10) Bert was wasting valuable time and money.

(11) Bert wasted valuable time and money.

The key to a solution to the imperfective puzzle is the observation that there is an important difference between the pair of sentences (8), (9) and (10), (11): in asking whether (8) $\models$ (9) one asks whether a culminating process entails its culmination; in asking whether (10) $\models$ (11) the question is essentially whether processes are downward persistent. To be precise, *Bert's writing a thesis* is classified as a culminating process, and the culmination *Bert wrote a thesis* is its completion. According to our BAF account there is no contradiction in continuations of culminating processes that explicitly deny its culmination:

(12) Bert was writing a thesis, but gave it up to join a heavy metal band.

Formally, in a sorted BAF failure of completion of a culminating process e is represented by the fact that the partial function `Compl` is not defined in e.

The above solves one half of the imperfective puzzle: (8) does not imply (9). How do we guarantee that (10) implies (11)? This is a simple consequence of clause 7 of the definition of a sorted BAF. Identifying *Bert's wasting* ... as a (non-culminating) process, we have for any sorted BAF $\mathbf{B}$, and any interval i in that BAF:

$$\begin{aligned}&\mathbf{B}, i \models \textsc{Past}\,\textsc{Prog}(\mathit{Bert}\ldots)\\ &\quad \text{iff}\quad \exists j\,(j < i \;\&\; \mathbf{B}, j \models \textsc{Prog}(\mathit{Bert}\ldots))\\ &\quad \text{iff}\quad \exists j, e\,(j < i \;\&\; e \in P_{\mathit{Bert}\ldots} \;\&\; \mathtt{Proc}(e) \;\&\; j \sqsubseteq^{+} z(e)).\end{aligned}$$

But this means that $j \sqsubseteq z(e)$, and hence $e\mathcal{Z}j$, and thus

$$\mathbf{B}, i \models \textsc{Past}\,(\mathit{Bert}\ldots),$$

and (10) implies (11).

6.4 Conclusion

In this extended abstract we have sketched, in very simple terms, how combined ontologies can be used in the semantics of temporal constructions. To conclude we briefly discuss our ongoing work on richer, more realistic systems, and note other BAF-like proposals we have found in the literature.

Sorted BAFs incorporate some of the Moens and Steedman ideas, but a great deal remains to be done. For example, although the sorts and the `GRiTo` and `Compl` relations model something of the Moens and Steedman notion of subevent structure, they don't capture the important idea that this subevent structure is recursively formed out of entities called nuclei. A nuclei is essentially a little 'package' consisting of a culminating process, a culmination, and a consequent state. Sometimes one wants to look at the internal structure of such packages, and sometimes one wants to treat this package simply as a 'lump' which can be linked to other packages. We are currently working with what we term *nucleic* BAFs. These are BAFs in which the eventuality occurrences are recursively generated out of Moens and Steedman style nuclei. Using such structures makes it possible to give analyses of a number of phenomena: in particular, we have given a Moens and Steedman style analysis of adverbial modification, and moreover can account for the interaction of progressive and perfective aspect in a natural way. (This is a topic that Moens and Steedman do not consider.) We are working on the semantics of temporal connectives (such as '*when*' and '*until*') in the setting of nucleic BAFs. An important part of this work is to reconstruct in the (essentially static) BAF framework an analogue of the (essentially dynamic) notion of 'type coercion' used by Moens and Steedman.

But these are topics for the full version of the paper. What can be said at a more general level concerning the idea of using combined ontologies in the study of temporal semantics?[2] We find the approach appealing for a number of reasons. First, it is intuitive. Pre-theoretical talk is often couched in terms of a mixture of different sorts of entities and their interrelations. Rather than ignore these intuitions, it seems better to try and be precise about them. Second, it seems to work. Formalisations couched in a single ontological setting tend to fare well with a handful of phenomena but can be extended only with difficulty. In contrast, we find the ease with which a wide range of phenomena can be modeled with BAFs striking. (We believe that most of the work of Moens and Steedman can be captured — and extended — in a manner that does no violence to its guiding intuitions.) Thirdly, the approach is, in a very useful sense of the word, conservative. It does not discard the work offered by point based, interval based or event based approaches: rather, it locates them in a richer setting. This retains what is good in earlier analyses, and lets the reasons for their shortcomings become clearly visible. To sum up, while BAFs as we have defined them here are only a crude approximation to the subtlety of temporal discourse, we feel that the underlying idea of combining ontologies will prove important.

To close the abstract we briefly note some other multiple ontology or BAF-like approaches we are familiar with. First, Verkuyl has long advo-

[2] Actually, the idea of combining ontologies seems of importance in many other areas of applied logic as well; see Blackburn and de Rijke (1994) for further discussion.

cated the use of such structures, and for a wide variety of reasons. For example, by using back-and-forth links between the real numbers and the natural numbers he is able to consider both discrete and continuous perspectives on a given event. For an overview of his work, see Verkuyl (1993). Next, building on the work of Verkuyl, Oversteegen (1989) analysed the semantics of various English and Dutch expressions in terms of certain moves between an 'objective' and a 'subjective' time flow. Although her structures differ from ours — the 'objective' flow is like an interval structure and the 'subjective' flow is a discrete time line — her approach has many ideas in common with ours. Tense, and perfective and progressive aspect are analysed in terms of a number of basic transition patterns between the structures. Her analysis of Dutch temporal constructions is quite detailed, and we think it would be interesting to formalise her discussion in terms of BAF-like structures.

Second a back-and-forth picture can be found in Seligman and ter Meulen (1992). This aspect of their work may not be immediately obvious, for most of their discussion is devoted to the construction of Dynamic Aspect Trees. Nonetheless, their idea of 'classifying interval frames' involves moving back-and-forth between two structures, and (we would argue) it is this that gives the needed flexibility to drive their dynamic system.

Lastly, our account seems to have affinities with Situation Semantics. This is clear if the Channel Theory initiated by Seligman (1990) is considered. In his terms we are using an interval structure to classify eventuality occurrences. Our treatment of the English present perfect essentially says that the peculiarities of the construction are due to the fact that it exploits this channel in a particularly strong way. More generally, Situation Semantics has long emphasized the importance of ontological diversity, and the way we evaluate formulas in BAFs could be regarded as an instance of their 'relational account' of meaning.

Acknowledgements

Patrick Blackburn and Maarten de Rijke would like to acknowledge the financial support of the Netherlands Organisation for Scientific Research (NWO), project NF 102/62-356 'Structural and Semantic Parallels in Natural Languages and Programming Languages'. Claire Gardent would like to acknowledge the financial support of the LRE 61-062 project 'Towards a Declarative Theory of Discourse'.

References

Benthem, J. van. 1991. *The Logic of Time.* Dordrecht: Kluwer Academic Publishers. second edition.

Blackburn, P., and M. de Rijke. 1994. Logical aspects of combined structures. In *This volume.*

Carlson, L. 1981. Aspect and quantification. In *Syntax and Semantics 14*, ed. P.J. Tedeschi and A. Zaenen. New York: Academic Press.

Dowty, D. 1979. *Word Meaning an Montague Grammar.* Dordrecht: Reidel.

Kamp, H., and U. Reyle. 1993. *From Discourse to Logic.* Dordrecht: Kluwer Academic Publishers.

Moens, M., and J. Steedman. 1986. Temporal information and natural language processing. Technical Report EUCCS/RP-2. University of Ediburgh.

Moens, M., and J. Steedman. 1988. Temporal Ontology and Temporal Reference. *Computational Linguistics* 14:15–28.

Oversteegen, L. 1989. *Two Track Theory of Time.* Doctoral dissertation, University of Amsterdam.

Reichenbach, H. 1947. *Elements of Symbolic Logic.* New York: Free Press.

Seligman, J. 1990. *Perspectives: a Relativistic Approach to the Theory of Information.* Doctoral dissertation, University of Edinburgh.

Seligman, J., and A. ter Meulen. 1992. Dynamic aspect trees. In *Logic at Work.* CCSOM, Amsterdam.

Vendler, Z. 1967. Verbs and Times. In *Linguistics and Philosophy, Chapter 4.* Ithaca: Cornell University Press.

Verkuyl, H. 1993. *A Theory of Aspectuality. The Interaction between Temporal and Atemporal Structure.* Cambridge: Cambridge University Press.

7

Naturalising Constraints

NICK BRAISBY AND RICHARD P. COOPER

Abstract

We examine the extent to which different formulations of conditional constraints respect the naturalisation requirement that they be grounded in non-intentional terms. This suggests an alternative formulation of constraints which we believe respects their motivating situation theoretic intuitions but meets the naturalisation requirement. In developing our alternative, naturalisation requires that we be explicit about the part-of relation that structures situations. This reduces to the vexed issue of persistence. We suggest a variety of treatments of non-persistent phenomena that allow us to assume infon persistence and an extensional treatment of part-of. This paves the way for a more naturalistic proposal concerning constraints—that they are propositional and 'borne' by situations.

Introduction

Within informational approaches to semantics (e.g., Dretske 1981, Barwise and Perry 1983), it is intended that semantic phenomena concerning, for example, linguistic meaning and the meaning of mental states, be given an explication in terms of physicalist descriptions of information flow. While such an assumption is argued to possess significant benefits, a difficulty arises. If an account of semantics is tied too closely to physicalist descriptions, its ability to account for error and mis-information appears compromised: cases of error are cases where a meaning-bearer does not signal its meaning, does not signal the physicalist state it conventionally signals. Thus, a major goal for any informational semantic theory is to provide for an adequate explication of both meaning and of error.

Barwise 1985 presents an account of natural language conditionals which

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

explicates the general fallibility of information flow. He argues that certain conditionals should be interpreted in terms of situation theoretic conditional constraints, constraints which permit the flow of information provided certain background conditions prevail. Barwise 1989a presents a different formulation of conditional constraints, one that is intended to remove some of the difficulties associated with background conditions. According to both accounts, the primary role of conditional constraints is as semantic and informational relations. Consequently, conditional constraints compete with other accounts of the fallibility of meaning relations, including, for example, Fodor's 1987 asymmetric dependence account.

Fodor's account is intended as an account of the meaning of mental states, such as that characterised by the thought 'lo, a cow.' The possibility of error then provides the following problem: a thought 'lo, a cow' may be caused, for instance, by a cunningly disguised horse, yet our thoughts about cows are seemingly just that—about cows, and not about disguised horses, or even cows or disguised horses (the disjunction problem: Fodor 1990: 59). How, then, is this property of aboutness robust to error? How are veridical thoughts to be distinguished from non-veridical thoughts in a non-question begging manner? Interpreted naturalistically, that they should be distinguished in a non-question begging manner means that they are to be distinguished in non-semantic, or non-intentional, terms (Fodor 1987: 98).

Our aim is to investigate whether different formulations of conditional constraints respect the requirement for naturalisation. We argue that no current situation theoretic formulation does so, but suggest an alternative account of constraints which we believe respects their motivating situation theoretic intuitions and naturalistic considerations.[1] In section 1 we briefly survey several situation theoretic accounts of conditional constraints, concentrating on two. The first assumes that constraints are both propositional and conditional while the second assumes that they are infonic and local. In section 2 we consider arguments concerning particular approaches to naturalisation and relate these to the previously articulated accounts of constraints. This discussion leads to the conclusion that the local account is to be preferred. However, grounding such an account requires that we be explicit about the part-of relation that may hold between situations. This reduces to the vexed issue of persistence, which we consider in section 3. We marshal arguments suggesting that constraints are local but propo-

[1]Our domain of interest in this paper clearly overlaps substantially with that of Koons 1994—both papers are concerned with how conditional constraints might admit error. However, we reject Koons' appeal to objective probabilistic relevance. Instead, we assume, more traditionally, that under ideal conditions information flow is deterministic, but that error arises from the application of constraints when they are not warranted.

sitional, and in section 4 we propose just such an account of conditional constraints, discussing its implications for the structure of situations, the part-of relation, infon containment and persistence.

7.1 Previous Accounts of Constraints

According to Barwise's 1985 analysis, conditional constraints hold provided certain background conditions prevail. For example, *ceteris paribus*, Claire's rubbing of her eyes means she is sleepy. But with pollen present, and irritating Claire's eyes, her eye-rubbing does not mean that she is sleepy: the relation between eye-rubbing and sleepiness is conditional. Following Barwise and Perry 1983, constraints are treated in terms of situation types. Positive conditional constraints are characterised as follows:

> If $S \Rightarrow S' \mid B$ holds, then if f is an anchor for the parameters in S, $s : S[f]$, $s : B[f]$, and s is actual, then there exists an actual situation s' such that $s' : S'[f]$.

There is an issue concerning the status of such constraints as infonic (i.e., supported by some situation) or propositional (i.e., situation independent). Though possible, taking constraints in this form to be infonic raises several problems. Firstly, they are intended to be downwardly (though not upwardly) persistent, at least in the sense of holding in a situation and all of its sub-situations, and thus contrast with infons as usually conceived, which are often regarded as being upwardly persistent, but never as being downwardly persistent. Secondly, as a background situation type is explicitly included as an element of the conditional constraint, it seems odd to regard such constraints as also being situated: it is unclear what would be purchased by the background situation type if one situation may support the constraint while some other situation does not. A propositional rendering of this approach to constraints seems most appropriate.

Barwise 1989a gives an alternative treatment of constraints where they are explicitly situated. In this account background situation types are not employed. Rather, constraints are infonic objects supported by some, presumably maximal, background situation. Involvement may be characterised as follows:

> If $b \models \langle\langle \textit{involves}, \vec{x}, \sigma, \tau; + \rangle\rangle$, then for each $s \trianglelefteq b$ such that s is actual and for each anchor f of $\vec{x}$ to $\mathit{Obj}(s)$ such that $s \models \sigma[f]$, there exists an actual situation s' such that $s' \models \tau[f]$.

Here, constraints are taken to relate infons rather than situation types. The explicit inclusion of a set of parameters as an argument of the constraint acknowledges the role which parameters must play in allowing constraints to hold across a variety of situations, but requires that, if constraints are to be non-parametric objects and truly the right sort of object to be supported by a situation, the relation within the infon must actually bind the parameters within its arguments: *involves* is thus no ordinary relation. Note also that it is the treatment of constraints as infonic, as explicitly supported (or not) by situations, which captures the intuition that the conditionality of conditional constraints is a matter of situatedness. This treatment of conditionality might more properly be termed locality: constraints, on this account, are localised, rather than merely conditional.

Seligman 1990 develops an account of conditional constraints which builds on the earlier accounts of Barwise. Seligman first defines the concept of a perspective—"a part of the world seen from one point of view" (Seligman 1990: 151)—and then relativises conditional constraints to perspectives. So a conditional constraint holds unconditionally within a particular perspective. In replacing Barwise's background situation types and background situations with his notion of a perspective, Seligman is suggesting that constraints are neither local nor conditional, but that they are relative. Relativity allows constraints to be local, but extends locality by allowing that they also be agent dependent.

Barwise 1991 and Barwise and Seligman 1994 have recently outlined an alternative formulation of constraints in terms of *indicating* and *signalling* relations. The idea is that constraints involve both a semantic relation between infons (the indicating relation), and a corresponding causal relation between situations which support those infons (the signalling relation). The signalling relation effectively grounds an instance of a constraint, associating the general relation between infons with particular situations. The account is intended to clarify the issue of which situation supports carried information (i.e., the identity of the situation s' in the above *involves* constraints) and to make clear the links between situation theory and domain theory.

Although Barwise and Seligman discuss the issue of conditional constraints, they do not relate their fallibility to other situation-theoretic constructs such as background conditions. As a consequence we do not regard these as more plausible accounts for naturalisation than previous formulations, and do not consider them in detail in this paper.

7.2 Naturalising Constraints

Informational semantic theories are thought to present a number of benefits principally through their adherence to naturalistic explications of meaning. Fodor 1990 offers two compelling examples. First, naturalisation involves an avoidance of intentional irrealism, the view that intentional terms are necessarily empty, devoid of reference. For an irrealist, talk of linguistic utterances and mental states as having meaning is a (perhaps understandable) mistake. Indeed, an intentional irrealist would presumably also adopt an eliminativist strategy with respect to scientific discourse, in which case the mistake is perhaps less understandable and of arguably greater import A benefit of naturalisation, then, is that if a naturalistic explication of meaning can be given, it can be demonstrated that intentional irrealism can be avoided and that our current scientific discourse has genuine objects of inquiry

Second, naturalisation involves an avoidance of semantic holism Holism is the view that the meaning of any one token within a language cannot be individuated in isolation from the meanings of all other tokens in that language Holism thus has devastating consequences for any attempts at semantic analysis and, as Fodor indicates, strikes also at our ability to offer psychological explanations Such explanations presuppose the ability to individuate mental states, typically via intentional terms such as "belief" If holism is correct, then the meaning of a mental state can only be individuated through individuating the meanings of all beliefs possessed Two people who offer the same belief report, for example 'I believe that the moon is made of cream cheese,' nonetheless possess different beliefs unless all of their other beliefs have the same content [2] Holism, then, would undermine our ability to isolate psychological laws, laws that quantify over appropriately individuated mental states

It is not hard to recognise the features that a naturalised account of meaning must have consider rigid designation the non-naturalistic theory of meaning offered by Kripke 1972 and Putnam 1975 This theory offers explications of meaning in terms of causal-historical chains of communication, by which the intentions of speakers to refer to an initially baptised kind are conveyed Uses of the sentence 'that is a cow' (pointing to a disguised horse) are then in error because the object in question is not the same as the kind originally baptised 'cow ' Thus, questions of meaning ultimately give rise to questions concerning naming and name-using practices However, the theory is non-naturalistic since talk of naming and name-using is intrinsically intentional Were the theory to relate talk of meaning to non-intentional talk it would be naturalised Thus, any naturalistic account of meaning must avoid such intentional talk

7.2.1 The Disjunction Problem

In order to account for the meaning of mental states (and mental state terms), as Fodor 1987, 1990 indicates, a theory of meaning has to satisfactorily account for cases of error Consider, for example, the tokening of

[2] Stich and Laurence 1994 argue that both intentional irrealism and holism may be avoided even if we are unable to furnish a naturalised account of meaning Nonetheless it appears to us that a naturalised account is required in order to *demonstrate* that intentional irrealism and holism are false doctrines

the thought 'lo, a cow.' This can be tokened in ways which respect the presumed causal relation between the thought and its content (a cow) but, also, in ways which do not. The thought may be tokened for example by a cunningly disguised horse, or a large dog on a dark night, and so on. Errors such as these present problems for any simple causal account of the meaning of mental states and the problem is compounded in the case of mental states because, in the domain of thought, we do not have the analogues of linguistic baptisms, name-using practices and the like. If such an account specifies that the meaning of the thought is determined by that which causes its tokening, then the thought can only mean 'lo, a cow or a cunningly disguised horse or ...'. This, then, is the disjunction problem (Fodor 1990: 59), and any account of meaning, if it addresses the meaning of mental states, must explain how it is to be avoided.

It might be objected that Fodor's exposition of the disjunction problem is too narrowly conceived. In particular, Barwise and Perry 1983 and Israel 1987 argue that meaning and information, and the disjunction problem must be understood as involving both a forward-looking and a backward-looking aspect. For example, a mental state has meaning inasmuch as it has a causal antecedent of an appropriate type (the backward-looking aspect), but also causal consequents of appropriate types (the forward-looking aspect). A consequence is that individuation of cases of error may require consideration of their causal consequents. Israel, for instance, develops a teleological approach to the disjunction problem in which reference is made to the evolutionary appropriateness and success of consequent behaviour as determining the meaning of mental states. There are several counter-arguments, only one of which we sketch. Concentration on the forward-looking aspects of meaning seems to invite an albeit limited behaviourism, in that, as Fodor argues, evolutionary appropriateness is not sufficiently fine-grained to individuate meaning. Two states with different meaning, for example 'there is a fly in front of me' and 'there is a small black dot in front of me,' may offer an organism equal selectional advantage in a world in which small black dots also happen to be flies. Behavioural and evolutionary success appear to depend on actuality; meaning depends also on counterfactuality.

The disjunction problem arises for two reasons. Firstly, the goal of naturalisation is to specify, in physical (non-semantic, non-intentional) terms, how it is that certain relations acquire a semantic status.[3] Secondly, certain entities having a semantic value may nonetheless be physically tokened by

[3]We note in passing if error is to be accounted for within Barwise's 1991 account, then it may likewise be susceptible to the disjunction problem since it also explicitly relates a semantic (indicating) relation to an underlying causal (signalling) relation. The difficulty, then, is to explain how the indicating relation depends upon the causal

something other than their semantic value. The problem then requires for its solution that we offer an explication of the relation between an entity and its semantic value in non-semantic terms.

The disjunction problem may, in principle, be circumvented in a number of ways. Fodor 1990 criticises both Dretske's 1981 learning-based account of error and Millikan's 1986 teleological account. While the details of these accounts are not of primary importance, what is of significance for situation theoretic formulations of conditional constraints is the generality of his critique. Concerning ways of avoiding the disjunction problem, he argues that "all the standard attempts ... distinguish between two types of situations, such that lawful covariation determines meaning in one type of situation [type 1] but not in the other [type 2]" (Fodor 1990: 60). Having made such a distinction between type 1 and type 2 situations, solving the disjunction problem then requires "a convincing—and, of course, naturalistic—explication of the type 1/type 2 distinction" (Fodor 1990: 61). The problem that Fodor identifies with this approach to the disjunction problem is the notable difficulty in providing such a naturalistic explication. Indeed, Boghossian 1991 suggests a number of reasons to believe that it may not be possible to specify naturalistic conditions for being a type 1 situation. Further, he argues that even if such a specification is possible, there is no way of recognising such conditions and that, as such, theories should avoid positing semantic relations which are dependent on type 1 situations.

7.2.2 Naturalising Constraints

Given these considerations, it is clear that the background condition formulation of conditional constraints suffers from Fodor's criticism. Indeed, it is widely acknowledged that the background conditions which play a role in this formulation are difficult and may, in principle, be impossible to specify. That this is a problem appears to be testified by Barwise 1989a who claims as a virtue of his later formulation of constraints that it does not contain the "somewhat mysterious, or at least hard to manage, background conditions" (Barwise 1989a: 276). The problem with being unable to specify background conditions either in practice or in principle is that the account will always fall foul of the claim that it is question-begging. That is, it may always be argued that any complete specification of these conditions will necessarily include conditions which are semantic or intentional in nature. Background conditions need to be specified in naturalistic terms and this cannot be demonstrated unless the conditions can be specified completely.

Nonetheless, the possibility remains open that conditional constraints may be naturalised according to Barwise's later formulation. According to

---

relation but only indicates (semantically) appropriate causes and not causes which are instances of semantic error.

Barwise 1989b that situations are informational is due to their being part of a larger situation (a background situation) which supports the relevant information-bearing constraint. Situations which are mis-informational with respect to this constraint are, conversely, not part of this larger situation. Consequently, the task of specifying naturalistic conditions for information flow becomes the task of specifying (naturalistically) the conditions under which one situation is part of another.[4] That is, naturalisation requires an explicit basis for the part-of relation.

However, Barwise's discussion (Barwise 1989b) indicates that the part-of relation may, as it stands, be as wholly mysterious as the problematic background conditions. He argues that it is necessary to give up the idea that the part-of relation can be thought of in terms of infon containment. Motivation for this comes from a number of different sources, but the principal reason is that several phenomena appear to argue for the non-persistence of infons. It is these phenomena, then, that provide motivation for the view that a situation does not necessarily support all the facts supported by its sub-situations. Hence, on this view, the part-of relation cannot be reduced to infon containment. It is, however, unclear how else the part-of relation may be explicated. Thus, suggesting that infons need not be persistent renders this formulation of conditional constraints also resistant to naturalisation. Given that we see no possibility of naturalising other extant formulations of conditional constraints, our enterprise requires us to reconsider the motivation for rejecting infon persistence. It is to this issue that we now turn.

7.3 Persistence

Barwise 1989b abandons infon persistence as a general principle of situation theory, citing the following reasons:

1. Negative statements such as "no-one is sleeping". This fact will not persist: a larger situation may include someone who *is* sleeping;
2. Perspectival facts such as "a is left-of b" and it's negation. By compatibility, there must be a situation larger than both of these situations, yet, by coherence, it cannot contain both an infon and it's dual.

The first of these reasons seems due more to quantification (in this case existential quantification) than negation. The positive universally quantified statement "everyone is sleeping" suffers the same problem with persistence. In any case, Barwise takes these observations to mean that, for certain infons, persistence is not a valid principle. Is there another way of accommo-

[4]This, of course, assumes that situations can be individuated/specified according to naturalistic criteria, a point which we return to later.

dating such observations? If there is, then we can maintain our *prima facie* reasons to think of the part-of relation in terms of infon containment.

7.3.1 Persistence and Quantification

A number of approaches to the problem of quantificational non-persistence have previously been suggested. Barwise 1989b and Devlin 1991 suggest that 'basic' infons should be distinguished from 'non-basic' infons, where non-basic infons are in some sense derived from basic infons via, for example, quantification over some argument role. They define bi-conditional relations between quantificational infons and their corresponding basic infons. Thus, under this approach if a situation supports an infon then it must also support any infon obtained by existentially quantifying over any argument role of that infon, and *vice versa.* One difficulty with this approach is that, as Robin Cooper 1991b notes, it leads to large situations. Part of the motivation behind situation-theoretic approaches to natural language is that situations are small. An agent is not required to have direct knowledge of the entire world and all the ways it might have been in order to successfully interpret an utterance. Robin Cooper's approach was thus to consider only a one-way relationship between quantificational infons and their basic counterparts.

Richard Cooper has argued for an approach whereby phenomena that involve non-persistence are treated in terms of propositions, rather than infons. On this approach, quantifiers are treated as two place types which hold between a restriction object type and a scope object type. A quantificational proposition is true if and only if the set-theoretic equivalent of the quantifier relation holds between the objects of the restriction type and the objects of the scope type.

We suggest that yet a third account might be given by distinguishing between supported information and *carried* information. In his 1989b treatment of constraints, Barwise distinguishes between situations supporting information and situations *carrying* information with respect to some set of constraints. Given this distinction, quantificational information might be carried by a situation in virtue of that situation supporting various (basic) infons, together with a constraint arising from the structural definition of the quantifier. Carried information need not be persistent. In a larger situation the constraint which justified the original information may not hold. On this approach the wealth of information carried by a situation arises from world knowledge and part-of relationships that hold between situations, together with a small "core" of supported infons.

7.3.2 Persistence and Perspectival Relativity

The treatment of perspectival information is more problematic. Barwise 1989b proposes that the different perspectives, say on the *left-of* relation, result from the fact that one of its arguments roles (one corresponding to a viewpoint) has been projected. Thus *left-of* is really a relation that holds between three arguments—an object, a reference point, and an observer—but the observer argument role is often projected away in a sub-situation.

A related suggestion builds on Perry's 1986 proposal that many utterances do not contain components corresponding to constituents of the proposition they convey. Such constituents he calls unarticulated and they open up the possibility that perspectival relations may be treated as involving an unarticulated constituent, possibly representing point of view, or perspective (similar to the analysis considered by Macken 1990). Thus, a relation such as *left-of* is a three-place relation but expressing propositions involving *left-of* need not always involve articulating the argument role for perspective or point of view.

Two further possibilities are as follows. Firstly, we might argue that the relationship between the unprojected and projected versions of a situation does not correspond to the part-of relation. If the two situations are individuated according to different schemes of individuation, then they will be incommensurate. Secondly, the contrastive information may be carried by different constraints which hold relative to different background situations. In the case of *left-of*, one situation might support the absolute position of the object and its reference point, but in virtue of being part-of different background situations (corresponding to different perspectives), it may carry different relative positional information between the two. (This view is implicit in Braisby's 1990 treatment of word meaning).

7.3.3 An Extensional Account of $\trianglelefteq$

In virtue of these considerations, we do not judge that apparent non-persistence is a stumbling block to an extensional account of the part-of relation. We define part-of ($\trianglelefteq$) as follows:

If s and s' are situations, then

$$s \trianglelefteq s' \text{ iff } \{\sigma \mid s \models \sigma\} \subseteq \{\sigma \mid s' \models \sigma\}$$

This extensional treatment does not compromise the position which Barwise 1989b adopts of situations being metaphysically prior to infons: with a fixed scheme of individuation, the infons supported by a sub-situation of any situation s will indeed be a subset of those supported by s itself. Note though that we cannot be sure that *any* subset of infons supported by a situation will itself be a situation. However, if it is, it will be a sub-situation of the larger situation.

This definition also raises the issue of situation identity: when are two situations the same? Can two instances of the same situation support different infons? Are two situations which support the same infons necessarily identical? There are two points to be made here. Firstly, one might note the category error in the second question. Situations are instances of situation types. On this reading, an instance of a situation is a nonsense. Secondly, note that if two situations s and s' support the same infons, then we will have $s \trianglelefteq s'$ and $s' \trianglelefteq s$. It is tempting to infer from this that s and s' are one and the same situation, but without further criteria for situation identity this inference is not justified.

7.3.4 The Persistence of Constraints

It appears that we are now in a position to offer a more naturalistic treatment of conditional constraints. There remains one outstanding issue: the persistence or otherwise of constraints themselves. Thus far, we have pursued the possibility that constraints are local and infonic, yet if we take constraints to be infonic, and infons to be persistent, then constraints must also be persistent.

Barwise 1985 suggests that constraints are "local" in that they are confined in their application to situations and their sub-situations. Locality, in this sense suggests a form of *downward* persistence. To illustrate this, and the difference between the two approaches, let us consider in further detail the treatment of conditionality of the information carried by Claire's rubbing of her eyes. In situations where pollen is absent, Claire's eye rubbing signals that she is tired, but the constraint breaks down in the presence of pollen.

Under the approach of Barwise 1985, the conditionality of this constraint might be treated by restricting the constraint's domain to those situations of the type $[\dot{s}|\dot{s} \models \langle\langle \text{present, pollen; } - \rangle\rangle]$, that is, situations which support the fact that pollen is not present. This is not quite right though, as if pollen is not present in one situation then it will not be present in any sub-situation. This is not captured by the above type unless we require that the infon $\langle\langle \text{present, pollen; } - \rangle\rangle$ be downwardly persistent. Whilst downward persistence might seem appropriate for negative infons, adopting such an approach would require that even the smallest situation support a vast number of negative infons. We might rectify this by replacing the above type with $[\dot{s}|\dot{s} \not\models \langle\langle \text{present, pollen; } + \rangle\rangle]$, which is downwardly persistent (given that $\langle\langle \text{present, pollen; } + \rangle\rangle$ is upwardly persistent). This approach thus captures the apparent downward persistence of constraints by specifying a downwardly persistent background situation type.

On the approach of Barwise 1989b, Claire's eye-rubbing is analysed in terms of a, presumably maximal, pollen-free situation. It is this background situation which must support the constraint. The issue of downward persistence *per se* is avoided by explicitly dictating that constraints hold in all sub-situations of the situations which support them. Crucially, this approach does not require that a situation support a constraint for that constraint to hold in that situation. If this was to be required, constraints would need to be downwardly persistent.

Clearly Barwise's 1989b situated infonic constraints cannot be upwardly persistent. Such a constraint holds in all sub-situations of that which supports it. If it were also upwardly persistent, then any actual constraint (i.e., any constraint supported by an actual situation) would, by the compatibility of actual situations, hold in all actual situations. The constraint would thus lose any sense of locality or conditionality. Alternatively, if a constraint is supported by some actual situation s and that constraint is conditional, then there must exist some actual situation s' where the constraint does not hold. By the principle of compatibility, there must exist a situation of which both s and s' is a part. If constraints are upwardly persistent then this situation would also support the constraint, but this cannot be the case: s' denies the background conditions of the constraint so no situation of which s' is a part can also support the constraint. Conditional constraints can thus not be upwardly persistent. As such, constraints are radically different to infons in the normal sense. In short, their persistence properties are in complete opposition to the persistence properties of infons as originally conceived.

One apparent avenue of escape from this question of constraint persistence, open to Barwise 1989b and Devlin 1991, is to treat constraints as non-basic infons. Given that, on the standard account, non-basic infons are not required to be persistent, such an approach would appear to save us from the problem of persistence and allow us to maintain constraints as infons. If the persistence properties of constraints were to provide the only argument against an infonic treatment then this would indeed be the case. However, there are further arguments against such a treatment of constraints.

7.4 An Alternative Account of Constraints

7.4.1 Constraints as Propositional

If we were to take constraints as being infonic, it would be unclear how a constraint such as $\langle\langle \mathit{involves}, S, S'; -\rangle\rangle$ should be interpreted, or how the relation within an infonic constraint could bind the parameters which must occur in its arguments. Furthermore, constraints govern the structure of situations, and, on the approach of Barwise 1989b, the structure of the sub-situations of any situation which supports them. If a constraint holds in a situation then this has consequences for the infons supported by that situation and, on the approach of Barwise 1989b, the infons supported by its sub-situations. Richard Cooper 1991a has argued that the propositional level is the appropriate place for such structural information.

The third argument might be countered by distinguishing basic and non-basic infons, and taking constraints to be non-basic infons, as discussed above. However, this does not counter the first two arguments. In sum, these arguments suggest that constraints would be better treated not in terms of infons, but in terms of propositions which impose structure on situations and whose persistence is governed by internal properties of situations (i.e., the infons directly supported by situations).

7.4.2 Constraints as Borne by Situations

What, then, is a conditional constraint? In the light of the above, we take conditional constraints to be propositional objects. They consist of a background situation, an antecedent situation type, a consequent situation type and a variable binding operator which binds variables in the antecedent and consequent situation types. To emphasis the situatedness of such constraints, we notate positive conditional constraints thus:[5]

$$b \mid S \Rightarrow S'$$

Such a constraint holds in a situation s if and only if $s \trianglelefteq b$. If this is the case, and the constraint is actual, and if there exists an anchor f for the parameters in S such that s is of type $S[f]$, then s will carry the information (with respect to the constraint) that there exists a situation s' of type $S'[f]$. In symbols:

If $b \mid S \Rightarrow S'$, $s \trianglelefteq b$, and there exists an anchor f such that $s : S[f]$, then $s \mathrel{||\!\!-} \exists s'(s' : S'[f])$

Preclusion may be analogously characterised:

If $b \mid S \perp S'$, $s \trianglelefteq b$, and there exists an anchor f such that $s : S[f]$, then $s \mathrel{||\!\!-} \forall s'(s' \not: S'[f])$

Note that carried information, on this account is propositional. This allows for cases where the situation types in question are not infon-based. As suggested above, quantificational information might be of this form. Precluded information need also not directly reduce to infon terms.

We also allow reflexive versions of the above constraints, where the antecedent and consequent situations are identical:

If $b \mid S \Rightarrow_r S'$, $s \trianglelefteq b$, and there exists an anchor f such that $s : S[f]$, then $s \mathrel{||\!\!-} (s : S'[f])$

---

[5]This notation leaves variable binding implicit.

If $b \mid S \perp_r S'$, $s \trianglelefteq b$, and there exists an anchor f such that $s : S[f]$, then $s \Vdash (s \not S'[f])$

That we distinguish between supported information and carried information means that reflexive constraints do not lead to vast situations.

Rather than viewing $(\cdot \mid \cdot \Rightarrow \cdot)$ as a variable-binding operator which forms a proposition when given appropriate arguments, we treat constraints formed with $(\cdot \Rightarrow \cdot)$ (as a variable binding operator) as a previously unacknowledged sort of situation-theoretic object in their own right. Such objects are situated, like infons, and so may form propositions with situations, but they are not supported, and they do not consist of a relation, an assignment of arguments to the argument roles of that relation, or a polarity. They are genuinely different objects that are associated with situations in a genuinely different way (hence the use of the symbol '$\mid$' to indicate this relation rather than '$\models$'). For want of a better term, we christen '$\mid$' "bears". The background situation thus *bears* the constraint which allows situations to *carry* information.

It should be clear that this account of conditional constraints is not incompatible with the view presented by Barwise 1991 of indicating and signalling relations, provided that those relations are, like constraints, borne by situations. Although our account does not attempt to explicate the causal relationship between situations related by an *involves* constraint, it may be extended to include such a signalling relation.

7.4.3 Error and Non-Persistence

Here, we return to the motivating issue of the possibility of error. Our treatment of error follows Barwise's later account. If a constraint is misapplied in a situation which is not a sub-situation of some larger situation bearing the constraint, then the constraint will be mis-informational. Thus, though we might confuse a cunningly disguised horse for a cow, this is not because our cow-type mental state means "cow, or cunningly disguised horse, or ...", but because the situation causing the tokening of 'cow' is not part of a situation which bears the constraint that holds between cows and 'cow' tokens. Thus, on this account, the distinction between information and mis-information reduces to the issue of when a situation is part-of another. This in turn reduces to the relationship between the infons they support. Hence, we have opened up the prospect that conditional constraints may be offered a naturalistic interpretation.

7.5 Conclusion

Constraints are ubiquitous things, but previous situation-theoretic accounts of them have left important questions unanswered. Most notably, the applicability of a constraint in a situation has never been precisely pinned down.

In Barwise's early treatment, the use of unspecified background situation types was acknowledged to be problematic, but Barwise's later treatment, in terms of the part-of relation, was equally mysterious. In essence, our proposal is that an extensional treatment of part-of clears up this mystery, but in justifying and presenting this we have needed to counter a number of arguments. We are left with a view of situations with two primitive relations (or binary types, to be more precise): the familiar supports relation, which may hold between situations and infons, and the novel *bears* relation. Thus, although our treatment is extensional, we do not reduce situations merely to sets of infons.

Certain notes of caution are in order. First, we noted earlier that a naturalistic treatment of constraints will ultimately require the provision of naturalistic criteria by which situations are to be individuated. At present, there appear to be relatively few concrete suggestions for what such criteria might be. Second, regarding situation theoretic objects, claims of a metaphysical nature have been made on what sometimes appear to be wholly mysterious intuitions. Judgements of part-of relations that obtain between situations is a case in point: on what grounds are these to be intuited when it is simultaneously argued that part-of be treated non-extensionally? Until these and other difficulties have been resolved we remain modest but enthusiastic about our own proposal.

References

Barwise, Jon. 1985. Conditionals and Conditional Information. Technical report. Center for the Study of Language and Information. Reprinted in J. Barwise 1989, *The Situation in Logic*, 97–135. CSLI Lecture Notes Number 17. Stanford, CA: CSLI Publications.

Barwise, Jon. 1989a. Notes on Branch Points in Situation Theory. In *The Situation in Logic*. 255–276. CSLI Lecture Notes Number 17. Stanford, CA: CSLI Publications.

Barwise, Jon. 1989b. Situations, Facts and True Propositions. In *The Situation in Logic*. 221–254. CSLI Lecture Notes Number 17. Stanford, CA: CSLI Publications.

Barwise, Jon. 1991. Information Links in Domain Theory. Technical report. Indiana University Logic Group. Preprint #IULG-91-7.

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes*. Cambridge, MA: MIT Press.

Barwise, Jon, and Jerry Seligman. 1994. The Rights and Wrongs of Natural Regularity. In *Perspectives in Philosophy, 8*, ed. J. Tomberlin. 331–364. Atascadero, California: Ridgeview.

Boghossian, Paul A. 1991. Naturalising Content. In *Meaning in Mind*, ed. B. Loewer and G. Rey. 65–86. Cambridge, MA: Basil Blackwell.

Braisby, Nick. 1990. Situation Word Meaning. In *Situation Theory and Its Applications, I*, ed. R. Cooper, K. Mukai, and J. Perry. 315–341. CSLI Lecture Notes Number 22. Stanford: CSLI Publications.

Cooper, Richard P. 1991a. Persistence and Structure Determination. In *Situation Theory and Its Applications, II*, ed. J. Barwise, J.M. Gawron, G. Plotkin, and S. Tutiya. 295–309. CSLI Lecture Notes Number 26. Stanford: CSLI Publications.

Cooper, Richard P. 1991b. Three Lectures on Situation Theoretic Grammar. In *Natural Language Processing: EAIA 90 Proceedings*, ed. M. Filgeiras, L. Damas, N. Moreira, and A.P. Tomás. 101–140. Lecture notes in Artificial Intelligence, no. 476. NY: Springer-Verlag.

Devlin, Keith. 1991. *Logic and Information.* Cambridge, UK: Cambridge University Press.

Dretske, Fred. 1981. *Knowledge and the Flow of Information.* Cambridge, MA: MIT Press.

Fodor, Jerry. 1987. *Psychosemantics.* Cambridge, MA: Bradford Books.

Fodor, Jerry. 1990. *A Theory of Content and Other Essays.* Cambridge, MA: Bradford Books.

Israel, David. 1987. The Role of Propositional Objects of Belief in Action. Technical report. Center for the Study of Language and Information, Stanford University. CSLI-87-72.

Koons, Robert. 1994. Information, Representation and the Possibility of Error. This volume.

Kripke, Saul. 1972. Naming and Necessity. In *Semantics of Natural Languages*, ed. D. Davidson and G. Harman. 254–355. Dordrecht: Reidel.

Macken, Elizabeth. 1990. Out of the Mouths of Babes. In *Situation Theory and Its Applications, I*, ed. R. Cooper, K. Mukai, and J. Perry. 433–448. CSLI Lecture Notes Number 22. Stanford: CSLI Publications.

Millikan, Ruth. 1986. Thoughts without Laws: Cognitive Science without Content. *Philosophical Review* 85:47–80.

Perry, John. 1986. Thought without Representation. *Supplementary Proceedings of the Aristotelian Society* 60:137–151.

Putnam, Hilary. 1975. The Meaning of 'Meaning'. In *Language, Mind and Knowledge*, ed. K. Gunderson. Minnesota Studies in the Philosophy of Science, Vol. 7, 131–193. Minneapolis: University of Minnesota Press.

Seligman, Jerry. 1990. Perspectives in Situation Theory. In *Situation Theory and Its Applications, I*, ed. R. Cooper, K. Mukai, and J. Perry. 147–191. CSLI Lecture Notes Number 22. Stanford: CSLI Publications.

Stich, Stephen, and Stephen Laurence. 1994. Intentionality and Naturalism. Technical report. Center for Cognitive Science, Rutgers University. RUCCS-TR-9.

8

Reflexivity and Belief *De Se*

KAREN LEIGH BROWN

8.1 Reflexivity

A sentence with a reflexive pronoun, such as (1) or (2), can be used to make two different kinds of claims.

(1) Bob scratched himself

(2) Venus outweighs itself

The first sentence can be used to claim that Bob Bob-scratched or that Bob self-scratched; the second to claim that Venus Venus-outweighs or that Venus self-outweighs. If the difference is not obvious, consider possible inferences from (1) together with the information that Joe did what Bob did. On the Bob Bob-scratched reading of (1) we can infer that Joe scratched Bob. On the self-scratched reading we get that Joe scratched Joe. Or consider the difference in what property is being attributed to Venus in (2). On the first reading it is the property of being heavier than Venus—a property which many things, the sun, for instance, have. On the self-outweighs reading, on the other hand, Venus is claimed to have a property which nothing can have. So, the semantic contents of the two claims must somehow be distinguished despite the fact that it appears that the two will be true in the very same possible worlds. Soames (1986, 1992), Salmon (1992a, 1992b), Cresswell and von Stechow (1982), Cresswell (1985) and others have argued that since, indeed, they must be true in all of the same circumstances, any semantics where contents are treated as worlds or circumstances will be inadequate. But this is simply not true. It misses the fact that the move from possible worlds to situations is not a simple donning of blinders which restrict our view to smaller total parts of worlds. It is instead a move to

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

partiality. This is what makes Situation Theory so useful in modeling the semantics of natural language.

In discourse we are always operating with partial information. The claim in Situation Semantics is that when we talk we are describing situations, or parts of the world as individuated by agents. As a conversation progresses we continue to add information about the described situation, never reaching total information, but developing a picture of a situation based on which situation types it has been claimed to support. This partial information is what we work with. Just as a situation described in discourse need not resolve all of the issues that might arise in it, the situation which serves as the circumstance of evaluation, the content, of a sentence may be partial. It might, for instance establish that Bob Bob-scratched and yet leave open the question of whether Bob self-scratched. In any total situation where the first of these is true, the second will be also, but the same does not hold in every partial situation. This is because the minimal situation that supports Bob's self-scratching must contain more information than the minimal situation that supports that Bob scratched Bob. It must settle the issue of reflexivity. To separate these two contents then, we need to locate the source of that reflexivity in the situation. The task that remains is not that of finding a way to separate the two contents—situations do that for us. The pressing problem is finding an acceptable way to connect that difference in the situations to a difference at some level of linguistic analysis. We need a theory of linguistic reflexivity that meshes with the ontology of reflexivity.

Accounts of reflexivity offered by both linguists (Jacobson 1991, Chierchia 1989) and philosophers (Salmon 1992a, Salmon 1992b) have had three features in common. First, they have been extended to handle the problem of *de se* belief. Sentences like (3) have both a *de re* and a *de se* reading.

(3) Joe$_i$ believes he$_i$ is in danger

Take a situation in which Joe is having dinner at a restaurant that has mirrors covering its walls. Looking across the room Joe watches a waiter set a saganaki on fire. While the dish is still blazing, the waiter loses his balance and the flaming contents of his tray slide towards a man sitting at another table with his back to the waiter. That man is Joe. And Joe believes of that man that he is in danger. If Joe has not figured out that he is looking at himself, this is a case of *de re* belief, but if his knowledge is such that his belief is going to make him hop out of his seat, that is belief *de se*. The different readings here are very similar to those available for (1) and (2). The *de re* reading parallels the reading on which Joe would scratch Bob, while *de se* belief is similar to the use of a reflexive where Joe would scratch himself. Consequently, accounts of linguistic reflexivity have been constructed so as to be generalizable to handle *de se* belief.

A second feature that many accounts of reflexivity have shared is the postulation of what I call polymorphs, or shape changing verbs. The semantic type of these verbs varies independently of the syntax. The idea is that perhaps a verb like *scratch* usually designates a two place relation, but sometimes designates a property, and perhaps a verb like *believe* usually takes a proposition as its complement, but now and then can take a property. This is done either by postulating two separate lexical entries for the verbs in question, or by assuming a function that transmogrifies verbs in general.

A third aspect of these accounts, which is more of a side effect of the polymorphs, is the introduction of type mismatches. Mismatches crop up in two places. First there is a rift between the syntax and the semantics. That is, while in the syntax a given verb may always look like all the other two place relations do, in the semantics it will sometimes be a two place relation and sometimes not. Jacobson (1991) acknowledges this kind of mismatch, calling it a necessary ill, but does not address the further mismatch this kind of approach to reflexivity generates between the semantics and the ontology of belief. The problem is again that while there is a split at the level of semantics, at the other level there is unity. Given the semantics proposed for a verb like *believe*, one would expect two different kinds of believing to exist. For instance, when I believe something about my shoes and something about myself, say that we are all in my office, a theory of belief that matched up with a semantics containing polymorphs would have to make the implausible claim that I am simultaneously experiencing two different kinds of belief. I am believing a proposition about my shoes and I am self-believing the in-my-office property. Of the three common features of accounts of reflexivity, the only one that I would like to preserve is the first.

In this paper I will pursue the line that the two types of claims made using reflexive pronouns, the indirectly, or *accidentally* reflexive Bob scratched Bob claims, and the directly, or *essentially* reflexive Bob self-scratched claims are in fact separable in terms of circumstances. I will attempt to separate utterances that give us accidentally reflexive information from those that provide essentially reflexive information, treating simple cases of sentences containing reflexive pronouns and cases of *de re* and *de se* belief in a unified way, and avoiding the introduction of type mismatches and polymorphs. I will begin by considering some of the linguistic forms available for making reports of these kinds. In particular I will look at the behavior of SE[1] anaphors in Dutch, first in general, and then in perception reports where their distribution seems to require additional explanation. The way of handling the directly and indirectly reflexive distinction in perception reports will then be seen to extend naturally to cover the phenomena of *de re* and *de se* belief.

8.2 Reflexive Marking and Reflexive Readings

In discussing the distribution of the Dutch SE anaphor *zich*, I will be assuming in large part the variety of the Binding Theory developed in Reinhart and Reuland 1993. Reinhart and Reuland postulate that reflexivity plays an important role in binding phenomena. In their theory we have a syntactic condition, Condition A, which states that a reflexive marked syntactic predicate is reflexive, and a semantic condition, Condition B, which states that a reflexive semantic predicate is reflexive marked. A syntactic predicate consists of a head and all of the syntactic arguments to which it assigns a theta-role or case, plus one external argument (thus it includes the subject of a clause or the specifier of an NP). The semantic predicate of a head is that head and all of its arguments at some level in the semantics. Condition B is claimed to hold at the "relevant" semantic level. I will take that to be the level of situation types. A predicate is said to be reflexive if it has two or more coindexed arguments. I will consider a predicate at the level of situation types to be reflexive for the purposes of the binding theory if it has linked coarguments. That is, if, in the semantics, the roles of two arguments of the same verb are required to have their roles filled by the same parameter, then that predicate is semantically reflexive. This basic account makes very clear, and substantially correct, predictions about the distribution of *zichzelf*, which, by virtue of its *zelf* morpheme, reflexive marks predicates, and of *zich*, which does not. Consider the example in (4).

(4) Max haat *zich/zichzelf
 Max hates himself

Here *zich* satisfies Condition A vacuously, since it does not reflexive mark the predicate, but it violates Condition B since *zich* and *Max* are coarguments of *haat*, but that predicate is not reflexive marked. Both conditions are satisfied by *zichzelf* since the predicate is both reflexive and reflexive marked. So *zich* should be unacceptable while *zichzelf* should be fine. The binding theory appears to work. If, however, we consider the following examples of perception reports, its application is less straightforward.

---

[1]SE anaphors are morphologically simple, non-reflexivizing, referentially dependent pronouns, like *zich*, as opposed to SELF anaphors which are morphologically complex, reflexivizing, referentially dependent pronouns, like *zichzelf*.

(5) Henk hoorde *zich/zichzelf
Henk heard himself

(6) Henk be-keek zich/zichzelf
Henk watched himself

(7) Henk hoorde zich/zichzelf zingen
Henk heard himself sing

The expected pattern turns up in (5), but in (6) and (7) we should expect *zich* to be ruled out. Recall that *zich* was bad in (4) because it left us with a reflexive, non reflexive marked predicate. There are two possible ways around the dilemma in (6) and (7). If we are to maintain this basic binding theory, we must either claim that in fact the predicate is reflexive marked, or that in fact it is not semantically reflexive. Reinhart and Reuland claim that in (7) no reflexive semantic predicate is formed since *zich* is inside of a small clause and so not a coargument of *Henk*[2]. For (6), however, they take the first option, claiming that *be-keek* is intrinsically reflexive and so invisibly reflexive marks its predicate.

Reinhart and Reuland's solution for (7) makes good sense, but the story they tell about (6) is both unconvincing and unappealing. Claiming that *be-keek* is intrinsically reflexive [3] conceals a problem rather than solving one. It is especially ill motivated, given that the diagnostics Reinhart and Reuland propose for independently identifying the intrinsically reflexive verbs fail to pick out verbs like *be-keek*. This verb is capable of having a direct object distinct in reference from its subject and also fails the nominalization test—i.e., *watching is fun* does not most readily refer to watching oneself. We have then no independent reason for claiming that *be-keek* reflexive marks its predicate. The move to rest the weight of the theory on the precariously founded notion of intrinsic reflexivity is very unsatisfying. Still, this theory has made a promising start.

8.3 A Proposal

I would like to take advantage of Reinhart and Reuland's insight that the binding theory as they present it leaves many occurrences of anaphors unregulated, and to claim further that their binding theory does not capture all there is to reflexivity. I will maintain that reflexive readings do not depend on the presence of semantically reflexive predicates in the sense

[2]The acceptability of *zichzelf* in (7) requires further explanation then, but I will not delve into that here.

[3]And, in fact, intrinsic reflexivity is claimed for a large and diverse collection of verbs along with *be-keek*. Each of these verbs is said to be doubly listed in the lexicon—once as an intrinsically reflexive verb and once without that property—making the proposed tests very difficult to conduct.

of Reinhart and Reuland 1993. I will take the availability of sloppy VP anaphora as diagnostic of essential reflexivity. So, for instance, a sentence like (8) has two readings—one strict, one sloppy.

(8) Curtis knows himself better than Ted does

If we take (8) to mean that Ted does not know himself very well, we are getting the sloppy reading. We are reading *Curtis knows himself* as an essentially reflexive claim. If, on the other hand, we understand it as a comment on how well Ted knows Curtis, it is the strict reading we are working with. *Curtis knows himself*, then, is only accidentally reflexive. I will assume that semantically, reflexivity consists in argument role linking. It is semantic role linking that is important, not coindexing in the syntax. On both readings of (8) *Curtis* will be coindexed with *himself*. The readings are only distinguished by the way the semantic argument roles are filled.

When semantic coarguments are linked, the predicate must be visibly reflexive marked in the syntax in accordance with Condition B. When semantic non-coarguments are linked however, no such constraint is placed on the predicate, yet reflexive readings are still attained, as evidenced by the availability of sloppy readings. So there is no "semantically reflexive predicate", but we do have semantic reflexivity. Notice that I am not describing cases of logophoricity, since it will be argument roles that are linked. They are non-coarguments, not non arguments. If we assume that syntactic considerations independent of the binding theory may restrict the possible linking of argument roles in the semantics, we can also readily explain the presence of sloppy anaphora and the absence of strict anaphora in (9) and (10), and the availability of both for the case in (11). Here I am adopting roughly the system in Gawron and Peters 1990.

If we know that Sam did what Henk and Joe did, we get the following inferences from the a to the b sentences.

(9) a. Henk be-keek zich
b. Sam be-keek zich
Sam watched himself (not Henk)

(10) a. Henk hoorde zich zingen
b. Sam hoorde zich zingen
Sam heard himself (not Henk) sing

(11) a. Joe$_i$ saw his$_i$ pants burn
b. Sam saw his own pants burn or Joe's pants burn

Consider (9) first. Unlike Reinhart and Reuland, who claim that in (9)a. we have a semantically reflexive predicate, reflexive marked by the "intrinsically reflexive" *be-keek*, I contend that instead there is no semantically reflexive predicate here at all. We do however, obtain the sloppy anaphora in (9)b., and this, I claim is the mark of reflexivity. So what is going on in (9)? Well, *be-* is an inchoative marker, so *be-keek* is an inchoative verb. I assume therefore that it incorporates its affected argument at LF. As a result, the semantics of (9)a. is as in (12).

(12) $s \models \langle\langle [x \mid \langle\langle \text{BE} - \text{KEEK},\ \text{x}\rangle\rangle]subj : \text{h}\ \text{HENK}\rangle\rangle$

Here there is only one argument role. As a result, there can be no semantically reflexive predicate, and the binding theory is satisfied. That is, in (9) we have neither reflexive marking nor a semantically reflexive predicate, and hence the binding theory has nothing to say. Also, it becomes apparent that the sloppy anaphora in (9) is not truly sloppy anaphora. It is nothing more than what we have in the inference from *Beth skates and Lisa does what Beth does*, to *Lisa skates.* It only looks like sloppy anaphora because at s-structure *zich* is not yet incorporated. In (10) on the other hand, we do indeed have sloppy anaphora and reflexivity. What we do not have, however, are coarguments. In (4) *zich* was ruled out, because it was a coargument of *Max.* In cases like (4), where what is perceived is an individual, syntactic reflexive marking is required, but in (10), where the object of perception is a complex scene, we have a small clause, so *Henk* and *zich* are not coarguments and reflexive marking is not required. What remains to be accounted for is why only sloppy anaphora is available from (10)a.. The explanation for this is found in the syntax The anaphor *zich* lacks ϕ-features[4]. Without them it is uninterpretable. In (10)a., since *zich* is in a small clause, it cannot get the ϕ-features it needs directly from *Henk.* Instead it has to move at LF, if it is to be interpreted. Given that the only c-commanding head with ϕ-features is AGR, *zich* must move to INFL at LF. There it picks up the ϕ-features of *Henk.* This attachment is reflected in the semantics as an obligatory linking of argument roles. In the situation type, *zich* must be assigned the parameter that fills the subject argument role of *hoorde.* So the content of (10)a. will be represented as in (13).

(13) $s \models \langle\langle [x \mid \langle\langle \text{HOORDE},\ \text{x}\ (\text{s}' \models \langle\langle \text{ZINGEN},\ \text{x}\rangle\rangle)\rangle\rangle]subj : \text{h}\ \text{HENK}\rangle\rangle$

Semantically *Henk* and *zich* are still not coarguments, but they are linked arguments. Anyone who does what Henk is claimed to do in (10)a., fills the argument roles filled by the parameter x in (13). As a result, only sloppy anaphora will be possible. The referential dependence of *zich* is sufficient to guarantee semantic argument linking, and so reflexivity. Reflexive marking is not necessary. This is borne out by the fact that substitution of *zichzelf* for *zich* in (10) has no semantic effect. Both sentences will have the semantic form of (13). The situations they describe will be of the exact same type. The sentence with *zichzelf* will simply be more emphatic. What then accounts for the availability of strict anaphora in (11)? The differ-

[4]It is, for instance, not specified for gender

ence is simply that *his*, unlike *zich*, does have ϕ-features. It therefore need not move out of its small clause at LF. It does not attach to INFL. The semantic linking is for that reason not restricted by syntactic considerations. Given that *Joe* and *his* are coindexed, we have two possible semantic representations, which are given in (14) and (15).

(14) $s \models \langle\langle [x \mid \langle\langle \text{SEES},\ \text{x}, \text{s}' \mid (\text{s}' \models \langle\langle \text{BURN},\ \text{x's}\ {}_{(r\models\langle\langle \text{MALE}\ \text{x}\rangle\rangle)}\text{PANTS}\rangle\rangle)\rangle\rangle]$
$\qquad\qquad subj : \text{j JOE}\rangle\rangle$

(15) $s \models \langle\langle [x \mid \langle\langle \text{SEES},\ \text{x}, \text{s}' \mid (\text{s}' \models \langle\langle \text{BURN},\ \text{j's}\ {}_{(r\models\langle\langle \text{MALE}\ \text{J}\rangle\rangle)}\text{PANTS}\rangle\rangle)\rangle\rangle]$
$\qquad\qquad subj : \text{j JOE}\rangle\rangle$

In (14), by assigning *his* the parameter of the subject argument role of *sees*, we obtain the directly reflexive reading which provides for sloppy anaphora. This is what is often called a "bound variable" use of the pronoun. In (15), we have the corresponding "referential" use of the pronoun. *His* is assigned the parameter that fills the subject argument role of *sees his pants burn*. This provides for the strict anaphora, wherein anyone who does what Joe does, sees Joe's pants burn. This is only accidentally or indirectly reflexive. That is, given the linking in (15), it is only an accident that (11) is reflexive. The situation described on this reading is not claimed to be of a reflexive type. With (14) and (15) we are talking about situations of different types. Plainly the contents of directly and indirectly reflexive reports can be distinguished by situations. Just as we can distinguish a buying event from a selling event in terms of the situation types they support, despite the fact that one never occurs without the other also taking place in the same total situation, we can distinguish a case such as in (14) from a case like (15). Partial situations give us the flexibility needed to do this, and strict and sloppy anaphora demonstrate that it is these partial situations that we deal with every day.

One may well wonder how we know of a particular utterance whether we have (14) or (15). Of course sometimes we do not know and indeed a lot of humor depends on that potential uncertainty. But generally we do. The ambiguity is resolved by the context, circumstances, or discourse situation. Essential reflexivity is information which is determined by use. It is an extra issue usually left open by the syntax, but settled by the context. The syntax may restrict the possibilities, but if it does not, the sole determiner of essential reflexivity is the context. Reflexivity does not inhere in the verb—we have no need for polymorphs. Reflexivity is found in the way the context connects the semantic roles.

8.4 Reflexivity and Belief *De Se*

The same sloppy readings that give evidence of reflexivity in perception reports with *zich* can be found with *de se* belief. Naturally then, one would like the account given above to be applicable to *de se* beliefs as well. This would mean treating *de se* belief without recourse to polymorphs. Instead, situations should serve to distinguish contents and linking of argument roles should establish reflexivity. In the case of *de re* and *de se* belief it is even clearer that situations will be sufficient to distinguish the contents of the two kinds of claims. As discussed above, a report like the one in (16) is ambiguous between a *de re* reading, in which Joe believes of someone that he is in danger and although that someone is in fact Joe, Joe does not realize it; and a *de se* reading in which Joe does realize that that someone is himself.

(16) Joe$_i$ believes he$_i$ is in danger

Consider again the case of Joe and the flaming saganaki. An important fact to notice about the *de re*/*de se* distinction is that the only way to bring out the difference is to talk about two different situations. While any total situation in which Joe has the *de se* belief will also be one in which he has the *de re* belief, the reverse does not hold. We do not even need to move to partial situations to capture this difference.

First consider how these two readings will be represented in a situation theoretic account. We can assign two possible interpretations to (16). The difference will amount to different linking of argument roles through parameters. In both we assign the parameter $\jmath$ to Joe. Then we have the choice of assigning to *he* the parameter that fills the subject argument role of the verb phrase *believes he is in danger* or of assigning *he* the parameter that fills the subject argument role of the verb *believes*. If we follow the first possibility, we get the interpretation in (17).

(17) $s \models \langle\langle [x \mid \langle\langle \text{BELIEVES, x}, (\text{s}' \models \langle\langle \text{IN DANGER, j}_{(\text{r}\models\langle\langle \text{MALE}_{\text{J}}\rangle\rangle)} \rangle\rangle)\rangle\rangle] subj : \text{j JOE} \rangle\rangle$

Here it just happens that *he* and *Joe* end up getting their roles filled with the same parameter. As a result, it says that Joe has the property of being an x who believes that Joe is in danger. This interpretation gives us the accidentally or indirectly reflexive *de re* reading and will provide for strict anaphora. Anyone who does what Joe does under this interpretation will believe that Joe is in danger.

The other possibility gives us the interpretation in (18).

(18) $s \models \langle\langle [x \mid \langle\langle \text{BELIEVES, x}, (\text{s}' \models \langle\langle \text{IN DANGER, x}_{(\text{r}\models\langle\langle \text{MALE}_{\text{x}}\rangle\rangle)} \rangle\rangle)\rangle\rangle] subj : \text{j JOE} \rangle\rangle$

Here the argument role of *he* has to be filled by whatever parameter fills *Joe*'s role—the two roles are linked. As a result, (18) says that Joe has the property of being an x who believes that x is in danger. Here we have, I believe, the reflexive, *de se* reading. On this interpretation, anyone who does what Joe does will believe himself to be in danger. So this gives us the sloppy anaphora which is characteristic of *de se* readings. The different readings then are simply captured by linking the argument roles differently. This process mirrors, as it should, the way reflexivity was captured above in perception reports.

8.5 The Shape of the Theory

These results are very similar to the results that can be obtained using the formulas proposed by individuals working within various other paradigms. The structure in (17) resembles early Transformational Grammar accounts involving copying. Similarly, Cresswell advocates that a sentence like *John loves himself* be treated as expressing the same proposition as *John loves John* expresses. This strategy, which Salmon calls the Simple Anaphor Theory, produces the same readings that interpretations like (17) are designed to handle. As Salmon (1992b) points out, however, the Simple Anaphor Theory is inadequate since it leaves out the element of reflexivity often present in such reports. Following earlier work by Geach and Reinhart, Salmon (Salmon 1992a) proposes that what is really called for here is a bound variable interpretation. On such a theory the content of *John loves himself* is $(\lambda x)[x \text{ loves } x](\text{John})$, much as in treatments within Montague Grammar. This gives us just the type of reading that (18) does. Salmon (1992a) also acknowledges that we may need to retain the Simple Anaphor Theory and supplement it with the Bound Variable Theory. They all look awfully similar.

They are however importantly different. The situation theoretic account avoids introducing the type mismatches that crop up on accounts where there are two different *believes* relations. Like Cresswell, Chierchia (1989) and Jacobson (1991) also treat *de se* belief as a relation to a property. On Jacobson's account, being an x who believes P(x) is equivalent to being an x who *z*-believes P. Still, a sentence embedded under *believe* is represented as a proposition in the semantics, just like any other clause, while a sentence embedded under *z-believe* is translated as a property. And, of course, when we come to evaluate her semantic representation in a situation, or even a world, the fit is again not quite right. The double layer of mismatches created by taking *de re* and *de se* belief to be two different kinds of belief, rather than two different possible ways of hooking up the roles in a belief situation of one basic type, argues against maintaining that sort of approach to the problem, and separates the situation theoretic account offered here from the previous theories. Being able to generate all the readings is not enough. The shape of the theory we end up with is just as important.

8.6 Conclusion

Given that the situation based approach readily provides both the simple *de re* and the stronger *de se* readings that license the strict and sloppy anaphora, and does so without introducing any poorly motivated ambiguity in the lexical item *believe*, it ought to be preferred to accounts whose success depends on there being such an ambiguity. The discussion of *zich* above showed two things. First, it showed that what is essential to reflexive readings is a linking of arguments in the semantics. That this notion equally well captures the reflexivity in *de se* belief indicates that this is on the right track. Second, it demonstrated that the claim that many verbs are intrinsically reflexive, invisibly reflexive marking their predicates, is not only *ad hoc*, but also unnecessary. There is reason to believe then that claims of two *believes*, one intrinsically reflexive, are similarly unneeded. Just as in the syntax we have no real cause for recourse to intrinsic reflexivity, in the semantics of *believe* we find no reason to postulate more than one *believe* relation. It is to the advantage of this approach that it obviates the call for such a stipulative and unintuitive solution.

Moreover, the situation theoretic account brings out one aspect of the statement in (16) that other accounts of reflexivity miss, and that is essential to a proper understanding of the relation between the two readings. In Situation Theory a sentence like (16) is taken to classify the described situation as being of a certain type. It is not just predicating this or that property of Joe in some world. It is about a situation. In both cases, where (17) or (18) is true, the situation described is structured in such a way that it will support the situation type supported in (17). A situation that supports the type in (18) has to settle one more issue, but that issue aside, the situation in (17) and the situation in (18) are claimed to be of the same type. This is as it should be, since the reading in (18) entails the reading in (17). It is however plainly not claimed that the situation in (17) is of the type in (18). The crucial reflexive issue is left open by the structure given to the situation in (17). This captures the fact that the reading in (17) does not entail the reading in (18). This fact is missed by the Simple Anaphor and Bound Variable theories, as it is by Chierchia's (1989) *believe-1* and *believe-2* theory and Jacobson's (1991) *believe* and *z-believe* account. In all of these theories there are two different *believes*. An individual stands in the *believe-1* (or *believe*) relation to a proposition and in the *believe-2* (or *z-believe*) relation to a property. The relationship between the two kinds of belief is this: For all x, P [believe-2′ (P)(x) $\leftrightarrow$ believe-1′ (P(x))(x)]. The two formulas are logically equivalent. The two readings however, as we have seen, are not strongly equivalent. This is a case where even in a total situation the two claims are truth conditionally separable When considering only cases like *scratch* and working with total situations or worlds, the distinction is easy enough to miss. But with *believe* there is a clear and crucial difference.

On this account, we avoid the problem of type mismatches created by structuring the objects of *de se* belief as properties. Also, by incorporating into the analysis the insight that belief reports are claims that a partial situation is of a certain type, we are able to correctly capture the relationship of the *de re* to the *de se* reports. In addition, by relying on argument role linking rather than on intrinsic reflexivity, we arrive at a general theory of reflexivity—one that extends from an account of reflexive perception reports to cover *de se* belief—and we escape stipulative solutions. After considering the distribution of *zich* in perception reports it is no surprise that *de se* belief is best modeled not as a relation to a property, but as a linking of argument roles in situation types. These types of situations, in turn, serve to distinguish the contents of *de re* belief from the contents of *de se* belief, just as they do the contents of accidentally and essentially reflexive perception reports. The moral of all this, I think, is that the problem of reflexivity, rather than demonstrating that we need to pull back from circumstances, shows that we need to rely on them more heavily.

References

Chierchia, Gennaro. 1989. Anaphora and Atitudes *De Se*. In *Semantics and Contextual Expression*. Dordrecht: Foris Publications.

Cresswell, Max. 1985. *Structured Meanings: The Semantics of Propositional Attitudes*. Cambridge, Massachuseetts: MIT Press.

Cresswell, Max, and A. von Stechow. 1982. *De Re* Belief Generalized. *Linguistics and Philosophy* 5.

Gawron, Jean Mark, and Stanley Peters. 1990. *Anaphora and Quantification in Situation Semantics*. CSLI Lecture Notes, Vol. 19. Stanford: CSLI Publications.

Jacobson, Pauline. 1991. Functional Questions, Paycheck Pronouns, and Variable Binding. Paper presented at the ASL/LSA Conference, University of California, Santa Cruz, July, 1991.

Reinhart, Tanya, and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24.

Salmon, Nathan. 1992a. Reflections on Reflexivity. *Linguistics and Philosophy* 15.

Salmon, Nathan. 1992b. Reflexivity. In *Propositions and Attitudes*, ed. Nathan Salmon and Scott Soames. Oxford: Oxford University Press.

Soames, Scott. 1986. Incomplete Definite Descriptions. *Notre Dame Journal of Formal Logic* 27.

Soames, Scott. 1992. Direct Reference, Propositional Attitudes, and Semantic Content. In *Propositions and Attitudes*, ed. Nathon Salmon and Scott Soames. Oxford: Oxford University Press.

9

A Channel-Theoretic Model for Conditional Logics

LAWRENCE CAVEDON

Introduction

In a series of recent papers (Seligman 1990, Barwise 1993, Barwise and Seligman 1992, Seligman and Barwise 1993), Barwise and Seligman (henceforth *B&S*) have developed a mathematical model of *information flow*. Their model makes use of the notion of *channels*, which classify connections between tokens by the type of information flow that occurs between those tokens. As such, channels can be seen as structured objects that support *conditional information*, i.e., information of the form "if $\mathcal{P}$ then $\mathcal{Q}$". This has led Barwise and Seligman (1992) to suggest that their theory of channels may form the basis of a (Situation Theoretic) semantics of conditional sentences, whereby a conditional sentence is seen as making an assertion about a channel. In this paper, I take up *B&S*'s suggestion and sketch out a semantics for conditionals based on the theory of channels. In particular, I focus on using the logic of channels to capture valid patterns of inference involving conditionals.

There are two main issues which a logic of conditionals must address. The first is the intensionality of the conditional operator—the truth of a conditional is only loosely linked to the truth of its constituents. The second issue is the invalidity of certain classically valid patterns of inference—in particular, Transitivity, Monotonicity and Contraposition. Traditional conditional logics (see (Nute 1980) for an overview) tackle these prob-

Most of the work described here was performed while the author was at the Centre for Cognitive Science, University of Edinburgh, and was partially supported by The British Council under the Commonwealth Scholarship and Fellowship Plan. This work has greatly benefited from discussions with Robin Cooper and Jerry Seligman

Logic, Language and Computation
Jerry Seligmann and Dag Westerståhl, eds

lems by the use of possible worlds and *nearest-worlds* selection functions. The approach described below has more in common with that of Barwise (1989). Barwise models conditionals with Situation Theoretic *conditional constraints*, which are themselves intensional entities. Furthermore, conditional constraints have associated with them background conditions, which Barwise uses to model the context-dependency of conditionals and thereby invalidate the problematic patterns of inference.

While much of the underlying motivation of the semantics described below is the same as Barwise's, the Channel Theoretic semantics can be seen as an improvement in two specific areas. Firstly, it makes use of a better model of conditional information-flow. Channel Theory can be seen as a formal model for conditional constraints, resolving many of the issues that were left unanswered by previous attempts at such models. Secondly, whereas Barwise associates a type with each conditional constraint, denoting explicitly the background conditions under which the constraint is reliable, the model below represents background assumptions *implicitly*—the assumptions of a channel C are captured by the way in which C is related to other channels. The non-representation of background conditions within a channel is an important facet of Channel Theory, and also adheres to the widely-held view that explicit specification of assumptions is inherently impossible.

After a very brief presentation of Channel Theory, I present a semantics for conditional sentences and then introduce the formal mechanism that allows the representation of the context dependency of conditionals. This mechanism, a hierarchy of channels, is used in the definition of a number of operations on channels, effectively defining a logic for conditionals. This logic invalidates unwanted instances of Transitivity, Monotonicity and Contraposition, and has many attractive features, especially from a Situation Theoretic point of view. By assuming that the channels of a hierarchy satisfy certain weak conditions, it is shown that the logic supports many of the patterns of inference of standard conditional logics.

9.1 Channel Theory

In this section, I review the basics of *B&S*'s theory of information flow: Channel Theory. The following exposition is, by necessity, extremely brief—the reader is encouraged to consult the series of papers by *B&S*.[1]

[1]The best exposition of the philosophical foundations of Channel Theory is probably given in (Barwise and Seligman 1992); the most complete presentation of the formal model is Seligman and Barwise 1993.

9.1.1 Regularities

At the heart of *B&S*'s theory of information flow is the assertion that *regularities* are part of the natural order of things. A regularity is basically a relationship between properties (I use the term loosely) that allows inference to take place—given an instantiation of the first property, one is led to infer an instantiation of the second. A regularity is not to be based on simple correlation between events of certain types, but on real causal connection of some form. *B&S* certainly make no reductionist claims for their theory—they have no intention of defining the notion of regularity via some more fundamental concept. Their concern is the definition of a naturalistic model of information-flow based on regularities as intensional entities.

To *B&S*, a crucial property is that regularities can be both reliable (i.e., useful for inference) while still fallible (i.e., may admit exceptions). Their theory accounts for this seemingly conflicting situation by relativising regularities in a way that Situation Theorists will sympathise with—a regularity is supported by a *channel,* which basically captures contextual aspects of that regularity. However, even within a given context, exceptions to a regularity can still occur. *B&S* resolve this problem by ensuring that the particulars of the theory play an important role in the theory—it is not just the regularity between types that counts, but the connections between individuals that are classified by those types, and the way in which these connections are classified as instances of the general regularity.

B&S's resulting *Channel Theory* is to be seen as a model of conditional information flow. A channel is an object that supports information flow of a particular sort. The basis of the claim that we can define a semantics for conditional sentences based on Channel Theory rests on the assertion that the semantics of a conditional involves an instance of a regularity—i.e., a conditional asserts that a certain regularity holds over a certain pair of individuals in a particular channel. This idea is spelled out in more detail below.

9.1.2 Classifications

A classification is a structure that carves up (part of) the world into *tokens* and assigns various *types* to these tokens.

Definition 1 A *classification* A is a structure $\langle tok(A), typ(A), :_A^+, :_A^- \rangle$ consisting of a set of *tokens* $tok(A)$, a set of *types* $typ(A)$ and relations $:_A^+, :_A^-$ on $tok(A) \times typ(A)$. For $t \in tok(A)$ and $\phi \in typ(A)$, we say t is *classified positively (resp., negatively) by* ϕ *in* A if $(t :_A^+ \phi)$ (resp., $(t :_A^- \phi)$) holds.[2]

The tokens of a classification may be objects, individuals, situations, or even other classifications, while the types are any properties appropriate for

[2] I will usually drop the subscripts when this causes no confusion.

classifying (along some dimension) these tokens. A classification can be seen as a representation of information of a particular sort. This representation is highly relativistic—intuitively, one could view a classification as being relativised to a specific agent, epistemic state, point of view, or perspective on the world.

Seligman and Barwise (1993) define a number of operations on classifications, allowing the definition of more complex structures that represent complex Austinian propositions.[3] For simplicity and brevity, I will not use any such operations in this abstract. However, I will assume throughout that the set of types of a classification A has internal structure: in particular, $typ(A)$ is closed under type-conjunction, type-disjunction, and type-negation operations, denoted $\wedge$, $\vee$, $\neg$ respectively. I assume these operations to satisfy certain expected properties. Since for each $\phi \in typ(A)$ there is a corresponding $\neg\phi \in typ(A)$ satisfying $t :^{+} \neg\phi$ iff $t :^{-} \phi$, I write $t :^{+} \phi$ simply as $t : \phi$ and $t :^{-} \phi$ as $t : \neg\phi$. The types of A are ordered under a *type-entailment* relation $\leq_A$ satisfying some obvious properties with respect to the type operations. Type-entailment constrains the classification relation in the following way: if $t : \phi$ and $\psi \leq_A \phi$ then $t : \psi$. There is also a *type-conflict* relation $\perp_A$ on $typ(A)$, involving the expected associated interaction with the classification relation and with type-entailment (examples: if $\phi \perp_A \psi$ then not both $(a : \phi)$ and $(a : \psi)$ hold; $\phi \perp_A \neg\phi$).[4]

9.1.3 Channels

A channel C is basically a link between two classifications A and B, licensing the flow of information of some particular sort. Formally, a channel is itself a classification—the tokens of C are *connections* (each linking a token of A to a token of B).[5]

[3] For example, a conjunction operation takes classifications A and B and returns a classification $A \times B$ whose tokens and types are ordered pairs, respectively, of tokens and types of A and B, and whose associated classification relations are defined so as to mimic proposition-conjunction.

[4] These concepts are more fully defined in (Cavedon 1995). See also (Seligman and Barwise 1993).

[5] Seligman and Barwise (1993) define a channel as consisting of a classification plus two functions that map the tokens and types to their "endpoints", i.e., tokens and types of the linked classifications. In general, one should not identify the tokens and types of a channel with the pairs of tokens and types from the linked classifications as I do here—for example, there may be two distinct connections involving the same tokens. This problem is amplified in the presence of serial composition, defined below. However, this abuse leads to significant simplification and abbreviation of the rest of the presentation.

Definition 2 Let A and B be classifications. A channel $\mathcal{C} : A \Rightarrow B$ linking A and B is a classification $\langle tok(\mathcal{C}), typ(\mathcal{C}), :\rangle$.[6] The tokens of $\mathcal{C}$ are *connections*, denoted $s \mapsto s'$, with $s \in tok(A)$ and $s' \in tok(B)$. The types of $\mathcal{C}$ are *constraints*, denoted $\phi \rightarrow \psi$, with $\phi \in typ(A)$ and $\psi \in typ(B)$.
Some notation (adapted from (Seligman and J.Barwise 1993)): I sometimes refer to $\mapsto$ and $\rightarrow$ as *signalling* and *indicating* relations, respectively. Given a constraint $\gamma = \phi \rightarrow \psi$, $ante(\gamma)$ denotes ϕ and $succ(\gamma)$ denotes ψ.

Intuitively, a channel $\mathcal{C} : A \Rightarrow B$ regulates the flow of information between A and B. There are several important aspects to the theory of channels that makes it a suitable foundation for conditional information flow, such as the contextuality introduced by relativising the information flow to a classification (i.e., the channel itself), and the role played by the tokens. The type-token distinction allows a characterisation of error in information-flow, by way of *exceptions* to general regularities. This is not so critical to the account of conditionals, but is central to a Channel Theoretic analysis of generics and defaults (see (Cavedon 1995)).

As with other classifications, there are various operations that can be applied to channels, resulting in more complex channels. Operations that are of particular interest are those that relate to information flow. Given channels $\mathcal{C}_1 : A \Rightarrow B$ and $\mathcal{C}_2 : B \Rightarrow C$, we can *serially compose* them to obtain a channel that classifies information flow from A to C.[7]

Definition 3 Given connections $s \mapsto s', s' \mapsto s''$ and constraints $\phi \rightarrow \psi$, $\psi \rightarrow \tau$, let $(s \mapsto s' \,; s' \mapsto s'')$ be the connection $s \mapsto s''$ and $(\phi \rightarrow \psi \,; \psi \rightarrow \tau)$ the constraint $\phi \rightarrow \tau$. The (standard) *serial composition* of channels $\mathcal{C}_1 : A \Rightarrow B$ and $\mathcal{C}_2 : B \Rightarrow C$, written $(\mathcal{C}_1 \,; \mathcal{C}_2)$ is the channel $\mathcal{C} : A \Rightarrow C$ such that

- $tok(\mathcal{C}) = \{(s \mapsto s' \,; s' \mapsto s'') \mid s \mapsto s' \in tok(\mathcal{C}_1) \text{ and } s' \mapsto s'' \in tok(\mathcal{C}_2)\}$;
- $typ(\mathcal{C}) = \{(\phi \rightarrow \psi \,; \psi \rightarrow \tau) \mid \phi \rightarrow \psi \in typ(\mathcal{C}_1) \text{ and } \psi \rightarrow \tau \in typ(\mathcal{C}_2)\}$;
- the classification relation of $\mathcal{C}$ is the smallest relation ':' such that

$$\begin{array}{c}(s \mapsto s' \,; s' \mapsto s'') : (\phi \rightarrow \psi \,; \psi \rightarrow \tau) \text{ in } (\mathcal{C}_1 \,; \mathcal{C}_2) \\ \text{if } s \mapsto s' : \phi \rightarrow \psi \text{ in } \mathcal{C}_1 \text{ and } s' \mapsto s'' : \psi \rightarrow \tau \text{ in } \mathcal{C}_2.\end{array}$$ [8]

Serial composition is the only channel operation I consider in this abstract. To provide a more complete set of inference patterns, other op-

[6]There should of course be two classification relations here; however, since I do not make use of the negative classification relation, I will ignore it.
[7]Serial composition is the Channel Theoretic concept supporting Dretske's (1981) *Xerox Principle*.
[8]This last case can be stated a lot more succinctly in the absence of the aforementioned conflation between a connection and the pair of tokens it connects.

erations are required, namely *parallel composition* and *contraposition* (see (Barwise 1993)).[9]

9.2 A Channel Theoretic Semantics for Conditionals

In this section, I outline a semantics for conditional sentences based on Channel Theory. Since my main interest is in developing a logic that captures valid patterns of inference for conditionals, the presentation is brief and the semantics rather simplistic.

In Situation Theory, the content of a declarative utterance is an Austinian proposition, e.g., the claim that a certain situation is of a certain type. Given channels as objects, this idea can be extended to conditional sentences: a conditional sentence is taken to be a particular type of declarative sentence, one that makes a claim about a channel $\mathcal{C}$. The claim that such a sentence makes about $\mathcal{C}$ concerns the internal structure of $\mathcal{C}$, namely, that it supports certain information flow. More precisely, a conditional sentence asserts that a certain channel $\mathcal{C} : A \Rightarrow B$ contains a connection $s \mapsto s'$ amongst its tokens and a constraint $\phi \rightarrow \psi$ amongst its types. Of course, the conditional in question could be a counterfactual. In this case, $(s : \phi)$ does not hold in A. However, the assertion that $s \mapsto s'$ is an instance of the regularity $\phi \rightarrow \psi$ in no way depends on whether or not $(s : \phi)$ holds.

Definition 4 There is a classification $\mathcal{K}$ that has channels as its tokens and conditional facts as its types. Conditional facts are written $\langle \Rightarrow, \Phi, \Psi \rangle$, where Φ and Ψ are Austinian propositions. For $\mathcal{C} \in tok(\mathcal{K})$, $(\mathcal{C} : \langle \Rightarrow, (s : \phi), (s' : \psi) \rangle)$ holds in $\mathcal{K}$ iff $s \mapsto s' \in tok(\mathcal{C})$ and $\phi \rightarrow \psi \in typ(\mathcal{C})$.

The fact that a channel $\mathcal{C}$ is of type $\langle \Rightarrow, \Phi, \Psi \rangle$ can be read as "if Φ holds, then it carries the information Ψ, via $\mathcal{C}$". It should be clear that this conditional information (i.e., that $\mathcal{C}$ is of a certain type) holds regardless of whether Φ (and therefore Ψ) itself holds in the pertinent classification—it simply depends on the connections and constraints contained in $\mathcal{C}$. Note that $\mathcal{K}$, being a classification, can be one of the classifications with which a given channel is concerned. This allows a natural interpretation of nested conditionals: the content of a nested conditional is a proposition concerning a channel $\mathcal{C} : A \Rightarrow \mathcal{K}$.

As a simple example, consider the following conditional sentence.

"If the doorbell is ringing, then there is someone on the porch."

This sentence asserts a proposition of the form

[9]Seligman and Barwise (1993) model the contrapositive information-flow of a channel $\mathcal{C}$ as negative information flow in $\mathcal{C}$ itself, making use of negative classification. I have contraposition as a separate operation so that I can modify it to account for background conditions, as is necessary for obtaining an adequate logic of conditionals.

$$(\mathcal{C} : \langle \Rightarrow , (b : ringing), (p : occupied)\rangle),$$

where b denotes the doorbell in question, the type *ringing* holds of b just in case b is ringing, p denotes the porch in question, and *occupied* holds of p just in case there is someone on p. This proposition holds whether or not $(b : ringing)$ holds. The important point here is that there is a regularity between ringing doorbells and occupied porches (i.e., there is a constraint $ringing \rightarrow occupied$ in $typ(\mathcal{C})$), and the particular connection between the doorbell and porch in question falls within the domain of this regularity (i.e., there is a connection $b \mapsto p$ in $tok(\mathcal{C})$).

Not all conditionals are such obvious instances of general regularities as the one above involving the doorbell. Of the general classes of conditionals distinguished by Pollock (1976), the Channel Theoretic analysis is best suited to *necessitation* conditionals—i.e., those involving a connection of some sort between antecedent and consequent. In particular, conditionals such as

"If the moon is made of green cheese then $2 + 2 = 4$"

are not supported, even though any conditional with a valid consequent is generally a theorem of possible-worlds conditional logics.[10] Although the requirement of a regularity underlying any conditional may seem a strong one, the fact that a regularity may be quite limited in scope[11] means that a wide variety of conditionals can be interpreted in a way that fits the Channel Theoretic view. The applicability of the Channel Theoretic model of conditionals is further discussed in (Cavedon 1995).

9.3 A Logic of Conditionals

9.3.1 The Problem with Serial Composition

An important feature of Channel Theory is the ability to serially compose two channels, such as $\mathcal{C}_1 : A \Rightarrow B$ and $\mathcal{C}_2 : B \Rightarrow C$, allowing the classification of information flow from A to C. For example, the doorbell channel $\mathcal{C}$ above can be seen as the composition of channels $\mathcal{C}_1$ and $\mathcal{C}_2$, where $\mathcal{C}_1$ supports facts of the form $\langle \Rightarrow , (b : ringing), (t : pressed)\rangle$ and $\mathcal{C}_2$ supports facts of the form $\langle \Rightarrow , (t : pressed), (p : occupied)\rangle$. (The token t denotes a particular doorbell-button; t is classified by the type *pressed* if t is being pressed.) The use of serial composition effectively gives us Transitivity, which is widely considered to be invalid for conditionals. A standard exam-

[10] Such conditionals are also invalidated by recent logics of conditionals based on Relevant Logic (e.g. (Hunter 1980), (Mares and Fuhrmann 1993)).

[11] This is the case if the context defined by the channel containing the regularity is very restrictive.

ple (which I will refer back to) demonstrating the invalidity of Transitivity for conditionals is the following:[12]

"If there is sugar and oil in the coffee, then there is sugar in the coffee."
"If there is sugar in the coffee, then it will taste good."

"If there is sugar and oil in the coffee, then it will taste good."

We clearly do not want to infer the third of these sentences from the first two. The problem in this example is that the second sentence involves an implicit assumption (namely, that no foul-tasting stuff is put into the coffee) which conflicts with the antecedent of the first sentence.

Barwise (1989) resolves the problem of Transitivity by associating a background type with each conditional constraint and limiting the applicability of serial composition to constraints with the same associated type. In Channel Theory, the background assumptions behind regularities are reflected in the tokens involved in connections in the pertinent channel. However, I argue that the token-level alone does not allow us to determine when Transitivity is inapplicable. For example, let $\mathcal{C} : A \Rightarrow B$ be the channel about which the conditional

"If there is sugar in the coffee, then it will taste good."

makes a claim. It will be the case that, for any cup of coffee c such that $c \mapsto c \in tok(\mathcal{C})$, c does not contain anything that would make it taste poorly. However, this (rather extensional) condition is not enough. For example, if c contains neither sugar nor oil (i.e., the above sentence is a counterfactual), then we expect to be able to assert the above conditional of it. Hence, we expect $c \mapsto c$ to be in $tok(\mathcal{C})$ in this case. But $c \mapsto c$ should clearly be contained in the logical channel $\mathcal{L}$ that supports the first sentence of the above example: if c were to contain both sugar and oil, then it would certainly contain sugar. Hence, using standard serial composition, the channel $(\mathcal{L}\,;\mathcal{C})$ supports the unwanted third sentence.

Restall (1995) has suggested that examples of the above sort can be explained by appealing to the way in which serial composition "filters out" signal-target pairs—in his view, the channel $\mathcal{C}$ obtained by the serial composition should contain no token-level connections. However, in Restall's account of counterfactuals (as in the one here), a counterfactual corresponds to a connection (in some channel) whose signal is not classified by the antecedent of the pertinent type-level regularity. This leads directly to the situation described in the previous paragraph. What seems to be the problem is that the background assumptions associated with the channel $\mathcal{C}$ cannot be fully expressed by the tokens of $\mathcal{C}$. While the assumptions are

[12]Actually, this example is generally used to demonstrate the invalidity of Monotonicity.

reflected in the tokens (i.e., no cup of coffee over which C is applicable contains oil), we need to reflect a stronger condition, one somehow involving a modality (i.e., no cup of coffee over which C is applicable *could possibly* contain oil).

The approach taken below to resolving this problem is similar to Barwise's (1989), in that a regularity's background assumptions are represented at the level of types. However, Barwise's method requires that *all* background assumptions are captured and represented by a type, a situation that is clearly unsatisfactory. In the next section, I describe a method for capturing background conditions *implicitly*—the assumptions associated with a channel C are not represented in C at all, but are captured in the relationship between C and other "more informative" channels.

9.3.2 Encoding Assumptions with a Channel Hierarchy

The following definitions specify orderings between classifications, constraints and channels. The intuition behind these definitions is to move to "more discriminatory" classifications/channels when moving upwards on the orderings. In the subconstraint relation, this involves adding extra conditions to a constraint, either by conjoining them to the antecedent or disjoining them to the consequent. The subchannel relation involves ensuring that each constraint in the lesser channel can be mapped to a more discriminatory constraint in the greater one. (The definition of the subchannel relation is illustrated by example in Figure 1.)

Definition 5 Classification[13] A is a *subclassification* of classification B, written $A \sqsubseteq B$, iff $tok(A) \subseteq tok(B)$, $typ(A) \subseteq typ(B)$, $\leq_B$ agrees with $\leq_A$ when restricted to $typ(A)$, and for each $a \in tok(A)$ and $\phi \in typ(A)$, $(a : \phi)$ holds in A iff $(a : \phi)$ holds in B.

Definition 6 Let $\phi \rightarrow \psi, \phi' \rightarrow \psi'$ be constraints contained in channels $C : A \Rightarrow B$ and $C' : A' \Rightarrow B'$, respectively. We write $\phi \rightarrow \psi \leq_{(A',B')} \phi' \rightarrow \psi'$ if:

1. $A \sqsubseteq A', B \sqsubseteq B'$,
2. $\phi \leq_{A'} \phi'$, and
3. $\psi' \leq_{B'} \psi$.

Definition 7 Let $C : A \Rightarrow B$ and $C' : A' \Rightarrow B'$ be channels, and f a function from $typ(C)$ to $typ(C')$. We say C is a $\leq_{(A',B')}$ *-subchannel* of C' *wrt*

[13] Recall that each classification A has a type-entailment ordering $\leq_A$ associated with its types—$\phi \leq_A \psi$ if ψ entails ϕ.

f, written $\mathcal{C} \sqsubseteq_{(f,A',B')} \mathcal{C}'$, if

1. $A \sqsubseteq A', B \sqsubseteq B'$;
2. $tok(\mathcal{C}) \subseteq tok(\mathcal{C}')$;
3. for all $c \in tok(\mathcal{C})$ and $\gamma \in typ(\mathcal{C})$, if $c : \gamma$ in $\mathcal{C}$ then $c : f(\gamma)$ in $\mathcal{C}'$;
4. for all $\gamma \in typ(\mathcal{C})$, $\gamma \leq_{(A',B')} f(\gamma)$.

We write $\mathcal{C} \sqsubseteq_{(A',B')} \mathcal{C}'$ if there exists an f such that $\mathcal{C} \sqsubseteq_{(f,A',B')} \mathcal{C}'$.

(For each of the above orderings, I drop the subscripts when this causes no confusion.) Points of interest regarding the "hierarchy" defined by a subchannel ordering include the following.

- The implicit assumptions of a channel $\mathcal{C}$ are captured by its position in a given subchannel hierarchy. The background conditions of a constraint $\phi \rightarrow \psi \in typ(\mathcal{C})$ are made explicit in some constraint $f(\phi \rightarrow \psi)$, for some f and $\mathcal{C}'$ such that $\mathcal{C} \sqsubseteq_f \mathcal{C}'$.[14]
- Not only is it the case that a channel's background assumptions are not explicitly represented within that channel itself, but there is also not necessarily *any* channel in a given hierarchy which contains *all* the background assumptions of a given channel $\mathcal{C}$. Different background assumptions associated with the constraints of $\mathcal{C}$ may be represented in different channels, with $\mathcal{C}$ being a subchannel of each of these.
- If we desire, we can view the use of a channel hierarchy as relativising the semantics of conditionals; i.e., if we change the hierarchy, we encode different background assumptions, and therefore modify the behaviour of the conditional logic. One possible way to think of this is to consider a channel hierarchy as being associated with a particular agent: the inferences licensed by the resulting conditional logic are then the assertible inferences given the agent's cognitive state.

For the sake of simplicity and brevity, I will assume the following for the rest of this abstract. These simplifications are not strictly necessary, and the second one is not made at all in (Cavedon 1995).[15]

1. I will often assume that the signalling relation of any channel $\mathcal{C}$ is "reflexive", in that if $s \mapsto s' \in tok(\mathcal{C})$ then $s = s'$.

[14] A similar idea is used by Wobcke (1989), who defines a semantics of conditionals using a hierarchy of situation types, where each such type corresponds to an AI *plan schema*. Elaborations of such schemata result in a hierarchy, which can be seen as capturing background assumptions of the schemata in a manner that is loosely related to the way this occurs in a channel hierarchy.

[15] The first assumption is useful when comparing the Channel Theoretic conditional logic to possible-worlds conditional logics.

2 I also assume that any background assumptions to a constraint are accounted for by a strengthening of the antecedent rather than a weakening of the consequent Hence, I assume that if $C \sqsubseteq_f C'$ and $f(\phi \rightarrow \psi) = \phi' \rightarrow \psi'$, then $\psi = \psi'$

A final concept we need is that of *global* type-conflict This is so we can meaningfully talk about conflicting types when those types are taken from different classifications Given the simplifying assumptions that signalling relations are reflexive and that complex Austinian propositions involve complex types and a single token, type conflict is a strong enough concept to capture the notion of conflicting background assumptions

Definition 8 We say $\phi \in typ(A)$ and $\psi \in typ(B)$ *globally conflict*, written $\phi \perp \psi$, if there exists a classification C such that $A \sqsubseteq C$, $B \sqsubseteq C$ and $\phi \perp_C \psi$

9.3.3 Conditional Channel Operations

I discussed earlier how standard serial composition leads to Transitivity as an effective rule of inference However, the subchannel ordering provides a method for encoding the background assumptions of a channel Channel operations that are suitable for a logic of conditionals must be modified so as to account for such assumptions The following definition does this for serial composition An adequate logic requires other operations, such as suitable versions of parallel composition and contraposition These are defined in (Cavedon 1995)

Definition 9 Suppose we have channels $C_1 \; A \Rightarrow B$ and $C_2 \; B \Rightarrow C$, with $\phi \rightarrow \psi \in typ(C_1)$ and $\psi \rightarrow \tau \in typ(C_2)$ The *conditional serial composition* of C_1 and C_2, denoted (C_1 , C_2), is the same as the standard serial composition of C_1 and C_2 except that $\phi \rightarrow \tau \in typ(C_1 , C_2)$ only if there do not exist C'_1, C'_2 such that

1 $C_1 \sqsubseteq_{f_1} C'_1$, $C_2 \sqsubseteq_{f_2} C'_2$, and

2 $ante(f_1(\phi \rightarrow \psi)) \perp ante(f_2(\psi \rightarrow \tau))$ [16]

It should be clear from this definition that conditional serial composition effectively involves checking that the background assumptions of two constraints are compatible before composing them As such, this notion

[16]Pollock (1976) points out that the use of Transitivity involving an entailment" conditional can be problematic for necessitation conditionals For example, consider the following

"If I press the button, the bell will ring "
"If the bell rings, then it exists "

"If I press the button, then the bell exists "

As it is defined here, the operation of conditional serial composition supports in ferences such as this one A modification that avoids such inferences is given in (Cavedon 1995) of composition is not a great deviation from that of Barwise (1989). The important difference is the way in which these background assumptions are represented—Barwise uses a type to capture *all* background assumptions associated with any given constraint, whereas the use of the subchannel relation has background assumptions distributed throughout a hierarchy of channels.

As a simple example of the use of conditional serial composition, consider again the sugar-in-the-coffee example. Since the absence of oil is a background assumption, we assume the existence of a channel $\mathcal{C}'$, such that $\mathcal{C} \sqsubseteq_f \mathcal{C}'$, and $f(sugar{\rightarrow}good) = (sugar \wedge \neg oil){\rightarrow}good$. The conditional composition of the logical channel $\mathcal{L}$ with the channel $\mathcal{C}$ then fails to support the unwanted conditional. This situation is illustrated in Figure 1.

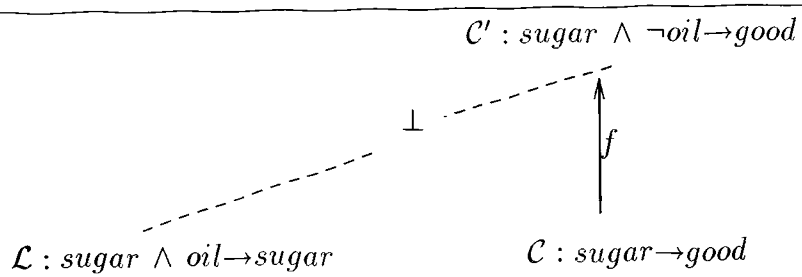


FIGURE 1 Failure of unwanted composition for the coffee example.

9.4 Adequacy of the System

I have shown how a channel hierarchy which encodes implicit background conditions can be used to define channel operations that invalidate the undesirable patterns of inference for conditionals. However, we also want to ensure that the resulting logic is sufficiently powerful, in that it supports inferences that are intuitively valid. In possible-worlds conditional logics, this can be done by imposing certain conditions on the nearest-worlds selection function. In the Channel Theoretic semantics, this is achieved by requiring candidate channel hierarchies to satisfy certain conditions.

9.4.1 Constraining the Channel Hierarchies

Some examples of possible constraints on channel hierarchies are defined below. An important aspect of these constraints is that they are independently motivated—i.e., they are *not* just conditions imposed on channels within a hierarchy so as to ensure certain patterns of conditional inference, but are conditions that all channels *should* satisfy, given *B&S*'s view of a channel. Other conditions are defined in (Cavedon 1995).

Reliability

The rationale behind the Reliability condition is that if a channel contains two constraints with the same antecedent, then these constraints should have mutually consistent background assumptions.

Definition 10 *Reliability Condition:*[17] Suppose we have a channel $\mathcal{C}$: $A \Rightarrow B$, containing constraints $\phi \rightarrow \psi$ and $\phi \rightarrow \tau$. Let $\mathcal{C}_1, \mathcal{C}_2$ be such that $\mathcal{C} \sqsubseteq_{f_1} \mathcal{C}_1$ and $\mathcal{C} \sqsubseteq_{f_2} \mathcal{C}_2$. Then it is not the case that

$$ante(f_1(\phi \rightarrow \psi)) \perp ante(f_2(\phi \rightarrow \tau))$$

This condition is clearly consistent with the view that a channel supports information-flow of a particular kind—if two constraints have such different background assumptions, then they should be in different channels. In fact, the following proposition shows that any channel that fails Reliability is inherently unreliable—i.e., its internal structure ensures that it licenses erroneous inference.

Proposition 1 *Suppose in the previous definition that $\mathcal{C}$ fails Reliability, and let $(a \mapsto b) \in tok(\mathcal{C})$ be such that $(a : \phi)$ holds in A Then $a \mapsto b$* must *be an exception to at least one of the constraints.*

Consequent Consistency

Under the assumption that signalling relations are reflexive, we need to ensure that the consequent of a given constraint does not conflict with the antecedent of that constraint, nor with its background assumptions.

Definition 11 *Consequent Consistency Condition.* Suppose we have a channel $\mathcal{C}$. $A \Rightarrow A$ with $\phi \rightarrow \psi \in typ(\mathcal{C})$. Then for any channel $\mathcal{C}'$ such that $\mathcal{C} \sqsubseteq_f \mathcal{C}'$, it is not the case that $ante(f(\phi \rightarrow \psi)) \perp \psi$.

As with Reliability, failure to satisfy the Consequent Consistency condition results in inherent unreliability.

Proposition 2 *Suppose in the previous definition that $\mathcal{C}$ fails Consequent Consistency, and let $(a \mapsto a) \in tok(\mathcal{C})$ be such that $(a : \phi)$ holds in A. Then $a \mapsto a$* must *be an exception to $\phi \rightarrow \psi$.*

9.4.2 Comparison to Standard Conditional Logics

Table 1 contains a number of axioms and rules of inference that are discussed in Nute's (1980) review of conditional logics.[18] An idea of the adequacy of the Channel Theoretic logic of conditionals can be obtained by

[17] (Cavedon 1995) contains a Reliability condition that is somewhat stronger than this one, but that satisfies the same properties

[18] For simplicity, I have omitted axioms and rules that are concerned with nested conditionals, modal operators, and rules of substitution. In the table, $\rightarrow$ is the conditional operator, $\supset$ the material implication operator, and $\Longleftrightarrow$ denotes logical equivalence

TABLE 1 Axioms and Rules of Inference discussed by Nute.

	Rules	*Supported?*
RCE	if $A \supset B$ then $A \to B$	yes
RCEA	if $A \iff B$ then $(C \to A) \iff (C \to B)$	yes
RCEC	if $A \iff B$ then $(A \to C) \iff (B \to C)$	yes
RCK	if $(A_1 \wedge \ldots \wedge A_n) \supset B$ then $((C \to A_1) \wedge \ldots \wedge (C \to A_n)) \supset (C \to B)$	yes
	Axioms	
MP	$(A \to B) \supset (A \supset B)$	yes
RT	$((A \wedge B) \to C) \supset ((A \to B) \supset (A \to C))$	yes
CA	$((A \to B) \wedge (C \to B)) \supset ((A \vee C) \to B)$	yes
CC	$((A \to B) \wedge (A \to C)) \supset (A \to (B \wedge C))$	yes
CEM	$(A \to B) \vee (A \to \neg B)$	no
CM	$(A \to (B \wedge C)) \supset ((A \to B) \wedge (A \to C))$	no
CS	$(A \wedge B) \supset (A \to B)$	no
CV	$((A \to B) \wedge \neg(A \to \neg C)) \supset ((A \wedge C) \to B)$	no
ID	$A \to A$	yes
MOD	$(\neg A \to A) \supset (B \to A)$	no
S*	$(\neg(A \wedge B) \to C) \supset ((\neg A \to C) \wedge (\neg B \to C))$	no
SDA	$((A \vee B) \to C) \supset ((A \to C) \wedge (B \to C))$	no
ST10	$(A \to B) \iff (\neg\neg A \to B)$	yes

determining which of these axioms and rules it supports.[19] Table 1 indicates which axioms and rules are supported by a Channel Theoretic logic of conditionals based on a channel hierarchy satisfying Reliability, Consequent Consistency, and the other constraints defined in (Cavedon 1995).

Clearly, any theory of conditionals that requires a connection between antecedents and consequents should not support axioms such as *CEM*, *MOD* and *CS*,[20] and these are supported by neither the Channel Theoretic logic nor the *relevant* logics of conditionals of Hunter (1980) and Mares and Fuhrmann (1993). Nute (1980) presents arguments as to why these axioms should be invalidated, and also argues against Lewis' (1973) *CV*, which is also not supported by the Channel Theoretic logic. However,

[19] A formal definition of the concept of axioms and rules being supported by the Channel Theoretic logic is far from straightforward, since an interpretation of the logical operations must be given—a major difficulty is that different instances of, say, the conjunction operator may correspond to different type-conjunction operations, in different classifications (even in the classification $\mathcal{K}$). See (Cavedon 1995).

[20] The first of these is the distinguishing axiom of Stalnaker's (1968) logic of conditionals, while the other two also hold in Lewis' (1973) logic.

RT (which defines a weakened form of Transitivity) is supported both by Nute and the Channel Theoretic logic.

The Channel Theoretic system supports many of the axioms argued for by Nute, such as *MP, CA, ID, CC* and *ST10.* One important difference is that Nute objects to the rule *RCEA*, mainly because support for this rule precludes support (in his logic) for S^* and *SDA* (which Nute argues should be validated)—Nute shows that, if the logical equivalence mentioned in *RCEA* is classical, then a conditional logical that contains both *RCEA* and *SDA* will also support Monotonicity. However, *RCEA* can be safely accepted in the presence of *SDA* and S^* if the logical equivalence is *relevant*. Moreover, Hunter (1980) gives an example that suggests that *SDA* is actually *not* an acceptable axiom. The other important axiom that Nute argues for, and which the Channel Theoretic logic does not support, is *CM.* For the Channel Theoretic system to support *CM,* it would require a property whereby every channel can be viewed as the parallel composition of two component channels.[21] While this would perhaps be a rather attractive perspective on Channel Theory—i.e., *any* channel can be decomposed into channels that focus on a particular component of the information-flow—I have not yet investigated it in any detail.

9.5 Discussion

Channel Theory seems to provide a solid explanatory basis for a number of tasks related to reasoning with conditional sentences, including the semantics of generics, default reasoning and AI task-planning. The use of Channel Theory as a formal basis for these tasks provides several important properties that address issues that have proved problematic to other approaches. In (Cavedon 1995), each of these tasks is modelled using the framework described in this abstract—i.e., the use of a hierarchy of channels to encode the implicit background assumptions of regularities. The use of a uniform framework in this way offers the possibility of providing important insights into the relationship between these different, yet clearly related, tasks. In particular, the move from a logic of conditionals to a system for reasoning with generics is very straightforward, with the resulting logic supporting many important patterns of inference (e.g., those specified by Asher and Morreau (1991)).

References

Asher, Nicholas, and Michael Morreau. 1991. Commonsense entailment: a modal theory of nonmonotonic reasoning. In *Proceedings of IJCAI-91.*

[21] Of course, a simpler condition could be imposed on channel hierarchies so as to validate *CM,* but I prefer to impose only conditions that should be expected of *any* channel.

Sydney.

Barwise, Jon. 1989. Conditionals and conditional information. In *The Situation in Logic*, ed. Jon Barwise. *CSLI Lecture Notes 17.*

Barwise, Jon. 1993. Constraints, Channels and the Flow of Information. In *Situation Theory and its Applications III*, ed. Peter Aczel, David Israel, Yasuhiro Katagiri, and Stanley Peters. *CSLI Lecture Notes 37.*

Barwise, Jon, and Jerry Seligman. 1992. The rights and wrongs of natural regularity. To appear in: James Tomberlin (ed.), *Philosophical Perspectives*, vol. 7 or 8. Available by ftp as `phil.indiana.edu:/pub/BarSel92.dvi`.

Cavedon, Lawrence. 1995. *A Channel Theoretic Approach to Conditional Reasoning.* Doctoral dissertation, Centre for Cognitive Science, University of Edinburgh.

Dretske, Fred. 1981. *Knowledge and the Flow of Information.* MIT Press, Cambridge, MA.

Hunter, Geoffrey. 1980. Conditionals, indicative and subjunctive. In *Papers on Language and Logic*, ed. John Dancey. Keele University Library.

Lewis, David. 1973. *Counterfactuals.* Harvard, Cambridge, MA.

Mares, Edwin D., and Andre Fuhrmann. 1993. A relevant theory of conditionals. To appear in *Journal of Philosophical Logic.*

Nute, Donald. 1980. *Topics in Conditional Logic.* Reidel, Dordrecht.

Pollock, John. 1976. *Subjunctive Reasoning.* Reidel, Dordrecht.

Restall, Greg A. 1995. Information flow and relevant logics. In *Conference on Information-Oriented Approaches to Logic, Language and Computation.* Moraga, CA.

Seligman, Jerry. 1990. *Perspectives: A Relativistic Approach to the Theory of Information.* Doctoral dissertation, Centre for Cognitive Science, University of Edinburgh.

Seligman, Jerry, and Jon Barwise. 1993. Channel Theory: Towards a Mathematics of Imperfect Information Flow (draft). Available by ftp as `phil.indiana.edu:/pub/SelBar93.ps`.

Stalnaker, Robert. 1968. A theory of conditionals. In *Studies in Logical Theory*, ed. Nicholas Rescher. Basil Blackwell, Oxford.

Wobcke, Wayne. 1989. A schema-based approach to understanding subjunctive conditionals. In *Proceedings of IJCAI-89.* Detroit, Michigan.

10

The Attitudes in Discourse Representation Theory and Situation Semantics

ROBIN COOPER

Introduction

Kamp 1990 is an important paper which emphasizes the role of information structure in an adequate treatment of the attitudes and presents this in terms of DRT. Kamp's paper presents the following major insights about the nature of the attitudes:

1. we need a theory of agents' mental states and their relationship to the world in order to be able to give an adequate treatment of the attitude reports. Kamp's paper concentrates on developing this more abstract theory.
2. we need to treat attitudinal objects in a bipartite fashion, representing both the internal state of the agent and the way that this is connected to the objects in the world external to the agent. In DRT this is represented in terms of externally anchored DRSs.
3. within the representation of the agent's internal state there needs to be a representation of background information which the agent might use to identify the objects of which the attitude holds. These are represented in Kamp's DRT by formal anchors within the DRS.
4. there need to be links between discourse referents used in the characterization of an agent's internal state and discourse referents in DRS used to make a (non-attitudinal) claim about the world. More importantly there need to be links between discourse referents used

Work on this paper was supported in part by Esprit BR 6852, DYANA-2.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

in the characterization of the internal states of agents in order to account for cases of intensional identity. That is, there can be links between two agents' mental states which nevertheless do not require that there actually be an object to which they are referring.

In this paper I will make the following claims concerning these points:

1. the aim of giving a general account of attitudinal states and not limiting ourselves to semantics for the attitudes is similar in spirit to the situation theoretic approach.
2. the bipartite proposal is essentially similar to proposals for treating the attitudes proposed by Barwise and Perry 1983, Barwise and Perry 1985. In fact it is quite surprising and encouraging that a straightforward recreation of Kamp's analysis in terms of situation theory yields a reasonably precise theory which is a recognizable variant of these proposals.
3. Kamp's notion of formal anchor seems best represented by the notion of restriction provided by situation theory. It is interesting that we can capture the notion of formal anchor within our framework using a theoretical tool that is independently needed.
4. the notion of linking can also be obtained rather straightforwardly by using the machinery of abstraction as developed in Barwise and Cooper 1991, Barwise and Cooper 1993 and Cooper 1992 based on work by Aczel and Lunnon 1991. The λ-calculus gives us a way of making roles in the characterization of mental states fall together by applying them both to assignments which provide new parameters which can be abstracted over. While this approach to sharing is not without its problems, it is not only pleasing in that we do not need to introduce new machinery specific to the attitudes, but also in that it gives us a particular view of what is meant by sharing discourse referents. It appears to give us one route towards a formal theory of mental states distributed over more that one agent. This topic will be taken up in a planned future version of the paper since it cannot be included here because of space limitations.

In addition I claim that the machinery we develop, coupled with the standard assumptions about the non-well founded nature of situation theory automatically gives us a treatment of mutual belief along the lines presented in Barwise 1989. This will also be taken up in a planned later version of the paper due to space limitations.

10.1 DRSs as Predicates

The central claim behind the reconstruction of DRT in situation theory is that DRSs are to be regarded as predicates, that is, either relations or types in the sense of Barwise and Cooper 1991, Barwise and Cooper 1993. Important here is Barwise and Cooper's use of Aczel-Lunnon abstraction to characterize predicates. There are two features of this kind of abstraction which are important for the reconstruction of DRSs as predicates.

simultaneous abstraction Any number of parameters in a parametric object may be abstracted over simultaneously. While in standard λ-notations one may have expressions such as

$$\lambda x, y, z[\phi(x, y, z)]$$

this is to be construed as an abbreviation for

$$\lambda x[\lambda y[\lambda z[\phi(x, y, z)]]]$$

In Aczel-Lunnon abstraction, however, it is the set which is abstracted over. Thus arguments to the abstract can be supplied simultaneously and there is no required order.

indexing This feature is closely related to the previous one. Since abstraction over parameters results in an object in which those parameters do not occur[1], we have to have some way of determining how arguments are to be assigned to the abstract in the case where more than one parameter has been abstracted over. Aczel and Lunnon achieve this by defining the abstraction operation in terms of indexed sets of parameters, i.e. one-one mappings from some domain ("the indices") to the parameters being abstracted over. An important aspect of this for us is that we can use any objects in the universe as the indices.

The leading idea is that we model discourse representation structures as abstracts which from the situation theoretic perspective are predicates. If we are working in situation theory, modelling DRSs as predicates gives us two options. They can be either relations or types. The choice is illustrated in (1) with respect to DRS corresponding to *a man owns a donkey* (ignoring matters of tense).

[1]This is important in order to achieve α-equivalence, i.e. $\lambda x[\phi(x)] = \lambda y[\phi(y)]$

(1) a. *Relation*

$$\begin{array}{|l|}\hline \quad i \to X,\, j \to Y \\ \hline \\ \text{man}(X) \\ \text{donkey}(Y) \\ \text{own}(X,Y) \\ \\ \hline \end{array}$$

b. *Type*

$$\begin{array}{|l|}\hline \quad i \to X,\, j \to Y,\, k \to S \\ \hline \boxed{S} \\ \\ \text{man}(X) \\ \text{donkey}(Y) \\ \text{own}(X,Y) \\ \\ \hline \end{array}$$

The difference is that in the relation there is no role for a situation whereas this is the case in the type. We will first consider the simpler option with DRSs as relations. We will then come to some motivation for considering them as types.

Note that whether we choose relations or types as the situation theoretic object to model DRSs we use roles of predicates (certain kinds of abstracts) to correspond to discourse referents. Thus we will talk of *discourse roles* rather than discourse referents, which represents a subtle shift in our view of what discourse representation is about.

10.1.1 DRSs as Relations

This means that we will take (1a) as the relation corresponding to the DRS for *a man owns a donkey.* How now do we get the effect of non-selective existential quantification that is obtained when traditional DRSs are interpreted in a model? We cannot use interpretation in a model since our DRSs are not syntactic objects. However, abstracts are the kind of thing that can be quantified over using a variant of the same technique that is used for the introduction of quantification into the λ-calculus as represented in (2).

(2) $\exists(\lambda x[\phi(x)])$

We introduce a distinguished property of relations "instantiated" or "realized" represented by $\exists$. This holds of a relation just in case there is some assignment to the roles of the relation which yields something that holds true when the relation is applied to the assignment. We can make this precise in situation theoretic terms as in (3).

(3) $$\begin{array}{|c|}\hline s \\ \hline \exists(r) \\ \hline\end{array}$$ is true implies

there is some assignment f appropriate to r such that

$$\begin{array}{|c|}\hline s \\ \hline rf \\ \hline\end{array}$$ is true

If desired (3) could be strengthened to a biconditional, though I do not believe that this is necessary.

This shows us that, given the relation corresponding to the DRS for *a man owns a donkey*, we can construct an infon where this relation is existentially quantified.

(4) $$\begin{array}{|c|}\hline \exists\left(\begin{array}{|l|}\hline i \to X, \jmath \to Y \\ \hline \text{man}(X) \\ \text{donkey}(Y) \\ \text{own}(X, Y) \\ \hline\end{array}\right) \\ \hline\end{array}$$

Notice the important effect of unselective binding here. Since we are using simultaneous abstraction we have simultaneous quantification. This is one important ingredient which enables us to capture the classical DRT analysis of donkey anaphora, though we will not have space in this paper to present the details.

Now that we have a quantified infon it is straightforward to use this to construct a proposition which might correspond to the interpretation of a DRS in classical DRT. If one does this in situation theoretic terms one such proposition is (5).

(5) $$\begin{array}{|l|}\hline \boxed{s} \\ \quad \exists\left(\begin{array}{|l|}\hline \quad i \to X, j \to Y \\ \hline \text{man}(X) \\ \text{donkey}(Y) \\ \text{own}(X,Y) \\ \\ \hline\end{array}\right) \\ \\ \hline\end{array}$$

But how are we going to achieve the effect of discourse anaphora if the DRS that you construct for a single sentence is modelled as an abstract where everything is already bound? It is here that the second feature of Aczel-Lunnon abstraction that I highlighted comes into play. The use of arbitrary role indices allows us to bind parameters but at the same time uniquely identify the roles in the abstract and identify roles across different abstracts. In designing grammars the strategy that I have been using for the incrementation of discourse representation is to assign a predicate corresponding to a DRS to each new sentence of the discourse and then integrate that predicate with the one obtained for the discourse so far. Basically the integration is predicate conjunction where roles that have the same index are merged. We define an operation of predicate conjunction, $\oplus$, which will be the central tool used in the incrementation of one DRS with another DRS (corresponding to the next sentence in the discourse).

The idea is best illustrated first by an example.

(6)
$$
\begin{array}{|l|}\hline \imath \to X,\ \jmath \to Y \\ \hline \\ r(X,Y) \\ \\ \hline \end{array}
\ \oplus\
\begin{array}{|l|}\hline \imath \to W,\ k \to Y \\ \hline \\ r'(W,Y) \\ \\ \hline \end{array}
$$

$$
= \begin{array}{|l|}\hline \imath \to X,\ \jmath \to Y,\ k \to Z \\ \hline
\begin{array}{|l|}\hline \imath \to X,\ \jmath \to Y \\ \hline \\ r(X,Y) \\ \\ \hline \end{array}
\begin{bmatrix} \imath \to X \\ \jmath \to Y \\ k \to Z \end{bmatrix} \\
\begin{array}{|l|}\hline \imath \to W,\ k \to Y \\ \hline \\ r'(W,Y) \\ \\ \hline \end{array}
\begin{bmatrix} \imath \to X \\ \jmath \to Y \\ k \to Z \end{bmatrix} \\ \hline
\end{array}
$$

$$
= \begin{array}{|l|}\hline \imath \to X,\ \jmath \to Y,\ k \to Z \\ \hline \\ r(X,Y) \\ r'(X,Z) \\ \\ \hline \end{array}
$$

In (6) we have two binary predicates which are conjoined by $\oplus$ to form a ternary predicate. The roles indexed by $\imath$ in the two original predicates are merged in the result If there had been no overlap in the indices the result would have been a quaternary predicate and if the roles of both binary predicates had been indexed by $\imath$ and $\jmath$ then the result would have been a binary predicate. Thus even though the parameters are bound, the indices are freely available and can be used to encode anaphoric relations. Note that it is important here that we are allowing arbitrary indices rather than, say, always using an intial segment of the natural numbers to do our indexing. It is the fact that we are allowed to use arbitrary indices which will give us the freedom to use them to encode discourse anaphoric relations.

Given the machinery for Aczel-Lunnon abstraction we have sketched here it is quite straightforward to give a general definition of $\oplus$.

(7) **Definition of $\oplus$**
If ζ is a predicate with role indices $r_{\zeta 1}, \ldots, r_{\zeta n}$, ξ is a predicate with role indices $r_{\xi 1}, \ldots, r_{\xi m}$ and f is an assignment whose domain $\{r_1, \ldots, r_k\} = \{r_{\zeta 1}, \ldots, r_{\zeta n}\} \cup \{r_{\xi 1}, \ldots, r_{\xi m}\}$ which assigns a unique parameter X_i (which is distinct from any free parameter in ζ or ξ) to each r_i in its domain then

$$\zeta \oplus \xi = \begin{array}{|l|}\hline r_1 \to X_1, \ldots, r_k \to X_k \\ \hline \zeta f \\ \xi f \\ \hline \end{array}$$

In Cooper 1993 I have characterized a basic DRT fragment based on part of the fragment defined in Kamp and Reyle 1993 including donkey anaphora, quantified sentences and relative clauses. In characterizing the fragment I exploit the fact that the use of Aczel-Lunnon abstraction allows us to use abstracts not only to recreate DRSs in the way I have presented here but also to combine that with the use of abstraction for compositional interpretation as in Montague's semantics.

10.1.2 DRSs as Types

The basic motivation for DRSs as types can be seen as soon as we look at DRSs of the kind which Kamp uses in his paper to introduce basic DRT concepts. Consider example (8) which he discusses there.

(8) Last month a whale was beached near San Diego. Three days later it was dead.

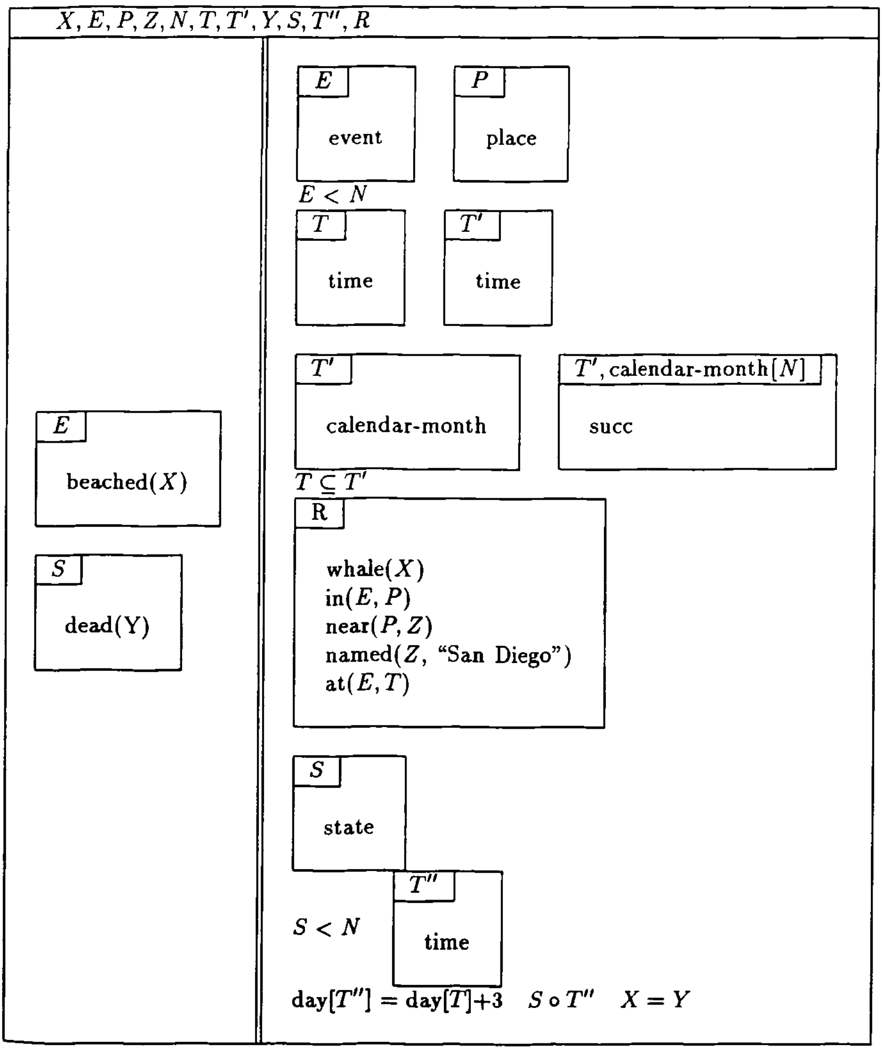
Kamp's DRS for the first sentence of (8) makes crucial use of a discourse referent for the event of the whale being beached and another discourse referent for the state of the whale being dead and there is a condition relating the temporal occurrence of the two. In his notation in the paper the relevant conditions look as in (9).

(9) a. $e \ldots \begin{array}{|l|}\hline \text{beached}(x) \\ \hline \end{array}$

b. $s \ldots \begin{array}{|l|}\hline \text{dead}(y) \\ \hline \end{array}$

Often in DRS notations ':' is used instead of '. . . '. In (10) I give a rough reconstruction of Kamp's DRS for the discourse as a situation theoretic type. I have made many arbitrary decisions here, for example, concerning which information is backgrounded as restrictions and the exact representation of temporal relations. My only aim here is to illustrate the relationship between DRT's ':' or '...' and the situation theoretic notion of a situation supporting an infon.

(10)



We can construct existential propositions from types in a similar manner as we did with relations.

(11) $$\begin{array}{|c|}\hline \begin{array}{|l|}\hline \quad X,Y \\ \hline \boxed{s} \\ \text{man}(X) \\ \text{donkey}(Y) \\ \text{own}(X,Y) \\ \hline \end{array} \\ \hline \exists \\ \hline \end{array}$$

Barwise and Cooper 1991, Barwise and Cooper 1993 adopt the following notation for objects such as (11).

(12) $$\exists X,Y \quad \begin{array}{|l|}\hline \boxed{s} \\ \text{man}(X) \\ \text{donkey}(Y) \\ \text{own}(X,Y) \\ \hline \end{array}$$

It is, however, important that we think of existential quantification as a predicate of predicates since this raises the possibility that we might construct propositions from DRSs that are other predicates of predicates for example *fictional* or, as becomes important when we talk of discourse referents linked between different agents, something like *classifies-mental-state*. Such predicates will not guarantee, of course, that the DRS is instantiated.

(13) a.

$$\begin{array}{|l|}\hline X, Y \\ \hline \boxed{s} \\ \text{man}(X) \\ \text{donkey}(Y) \\ \text{own}(X,Y) \\ \hline \text{fictional} \\ \hline \end{array}$$

b.

$$\begin{array}{|l|}\hline X, Y \\ \hline \text{“}a \text{ believes } X \text{ likes } Y\text{”} \\ \hline \text{classifies-mental-state} \\ \hline \end{array}$$

10.2 Externally Anchored DRSs

Kamp discusses the need for providing anchors for DRSs. The example he introduces is (14)

(14) That man is a cocaine dealer

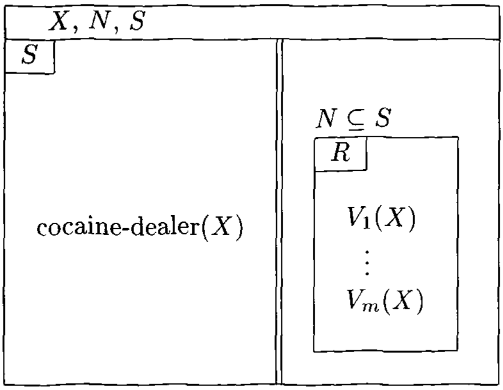
External anchors can be modelled in the situation theory of Barwise and Cooper 1991 as assignments to role indices.

(15) $$< \begin{array}{|l||l|}\hline \multicolumn{2}{|c|}{1 \to X,\ 2 \to N,\ 3 \to S} \\ \hline \boxed{S} & \\ \text{cocaine-dealer}(X) & N \subseteq S \\ \hline \end{array}, \begin{bmatrix} 1 \to a \\ 2 \to t \end{bmatrix} >$$

10.3 Formally Anchored DRSs

Kamp introduces the notion of a formal anchor in a DRS as a way of representing background beliefs which an agent might use to identify the objects which might be assigned to discourse referents. These can be modelled in situation theory as restrictions in the sense of Barwise and Cooper 1991.

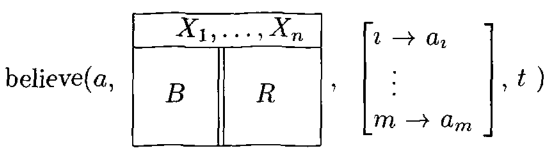
(16)



10.4 Mental States

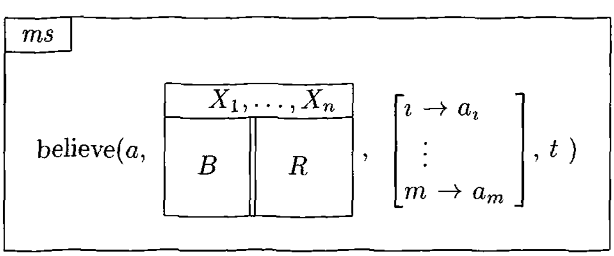
Putting the two kinds of anchors together we have the view of belief represented by the infon in (17) where the belief relation holds between an individual, a type (representing the internal state of the agent), an assignment (representing the external connections of the internal state to the environment) and a time (at which the belief is held).

(17)



This view is essentially a modernized version of the proposal made by Barwise and Perry 1983, Barwise and Perry 1985. It is developed further and related to other proposals in the literature by Cooper and Ginzburg 1994. (17) is to be regarded as an infon which is supported by a situation which we will call a mental state. Thus we will have propositions such as (18) where *ms* is a mental state.

(18)



10.5 Conclusion

In this paper we have shown how discourse representation can be modelled in terms of situation theoretic objects and how this leads us to an account of the attitudes which points to the close relationship between proposals in situation semantics and in discourse representation theory. This seems to be a promising line of research not only because it points to parallels in apparently diverging theories but because the two approaches to the attitudes have contributions to make to each other. The linking of discourse roles is something that has been discussed previously in DRT but not in situation semantics. On the other hand the use of situation theoretic objects to represent the objects of attitudes seems more attractive than the essentially syntactic analysis that is suggested by the discourse theoretical approach. Also the fact that situation theory takes into account non-well-founded objects offers the promise of combining the approach taken here with an account of mutual belief.

References

Aczel, Peter, and Rachel Lunnon. 1991. Universes and Parameters. In *Situation Theory and its Applications, Vol. 2*, ed. Gordon Plotkin Jon Barwise, Jean Mark Gawron and Syun Tutiya. CSLI.

Barwise, Jon. 1989. *The Situation in Logic.* Centre for the Study of Language and Information, Stanford University, CA. CSLI Lecture Notes Number 14.

Barwise, Jon, and Robin Cooper 1991. Simple Situation Theory and its Graphical Representation. In *Partial and Dynamic Semantics III, DYANA deliverable R2.1C*, ed. Jerry Seligman. Centre for Cognitive Science, University of Edinburgh.

Barwise, Jon, and Robin Cooper. 1993. Extended Kamp Notation: a Graphical Notation for Situation Theory. In *Situation Theory and its Applications, Vol. 3*, ed. Peter Aczel. David Israel, Yasuhiro Katagiri, and Stanley Peters. CSLI.

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes.* Cambridge, MA: MIT Press.

Barwise, Jon, and John Perry. 1985. Shifting Situations and Shaken Attitudes. *Linguistics and Philosophy* 8(1):105–61.

Cooper, Robin. 1992. A working person's guide to situation theory. In *Semantic Representation and Interpretation*, ed. Steffen Leo Hansen and Finn Sørensen. Frederiksberg: Samfundslitteratur.

Cooper, Robin. 1993. Towards a General Semantic Framework. In *Integrating Semantic Theories, Deliverable R2.1.A, DYANA-2*, ed. Robin Cooper. ILLC/Department of Philosophy: University of Amsterdam.

Cooper, Robin, and Jonathan Ginzburg. 1994. A Compositional Situation Semantics for Attitude Reports. In *Proceedings of the Fourth Conference* *on Situation Theory and its Applications*, ed. Jerry Seligman and ????

Kamp, Hans. 1990. Prolegomena to a Structural Theory of Belief and Other Attitudes. In *Propositional Attitudes: The Role of Content in Logic, Language and Mind*, ed. C. Anthony Anderson and Joseph Owens. CSLI.

Kamp, Hans, and Uwe Reyle. 1993. *From Discourse to Logic: Introduction to Model Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer.

11

A Compositional Situation Semantics for Attitude Reports

ROBIN COOPER AND JONATHAN GINZBURG

Introduction

There is agreement among a number of researchers that the attitudes should be analyzed in terms of structured objects which are fine-grained enough to prevent some of the troublesome inferences that arise from the classical possible worlds approach as represented by Montague 1974 (PTQ). It seems to us that the approach developed in Barwise and Perry is essentially similar in important respects to that developed within DRT[1] (Kamp, Asher, Zeevat) and also to that developed in the philosophical literature by Crimmins, Forbes, Richard. While, of course, there are differences in the various proposals (see, for example, the recent debate in *Linguistics and Philosophy* between Crimmins, Richard and Saul) there is something that all these approaches have in common: namely a concern with a structured analysis of mental states.

In this paper we shall (in section 11.1) recast Barwise and Perry's original ideas (Barwise and Perry 1983, Barwise and Perry 1985) using the kind of situation theory developed in Barwise and Cooper 1991. In section 11.2 we address the issue of defining a compositional fragment for the attitudes on the basis of the fragment defined in appendix 11.3. In an extended version of this paper, we will illustrate the account in greater detail and discuss the predictions concerning the fine grain of belief reports and show how to deal with embedded beliefs and quantifier scoping.

This research was supported in part by ESPRIT BR 6852 (DYANA-2).

[1]For discussion see Cooper's paper in this volume.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

11.1 Mental States

11.1.1 Preliminary Assumptions

A number of people have presented a triadic theory of belief, that is, one that treats belief not as a relation between agents and propositions but as a relation between an agent and two arguments in place of the proposition. The variant we present here is a reconstruction and elaboration of a variant of the triadic view presented by Barwise and Perry (Barwise and Perry 1983, Barwise and Perry 1985).

Ignoring the time argument, the standard Fregean view of the attitudes is "dyadic", i.e. attitude relations are seen as a relation between agents and propositions:

$$\alpha(a, p)$$

The important part of the triadic theory we characterize here is the use of two arguments ty and f in place of a single propositional argument as in the standard Fregean theory of the attitudes. The idea is that the type ty classifies the internal state of the agent. It corresponds to what Barwise and Perry called a "frame of mind". It comes along with various roles which may be linked to the world external to the agent. As the assignment is only partial there may be certain roles in the internal state which are not linked to objects in the world. In the case where there is a complete assignment, however, the result of applying the ty to f gives us a proposition, the "content"of the mental state. The intuition is that this is the same proposition as would be given on the dyadic view. The relation is expressed below:

$$\alpha(a, p)$$

$$\alpha(a, ty, f)$$

$$ty\ f = p$$

Thus the object of belief on the dyadic view is what we would call the content, the result of merging the internal state with the way that internal state is associated with objects in the world.

We treat mental states as a particular subclass of situations. We present some axioms that might be included in a theory of mental states below. We define notions of *mental state*, *content*, *exportation* and, *rationality*, as well as allowing for the possibility that different attitudes might allow for different reasoning about mental states.

Mental state

1. A *mental state* is a situation ms such that a proposition of the following kind is true:

$$\begin{array}{|l|}\hline ms \\ \hline \alpha_1(a_1, ty_1, f_1, t_1) \\ \dots \\ \alpha_n(a_n, ty_n, f_n, t_n) \\ \hline \end{array}$$

where α_i are *internal* attitude relations corresponding to *believe, know, desire*, a_i are agents, ty_i are types (possibly zero-place, i.e. propositions), f_i are partial assignments appropriate to ty_i, and t_i are times.[2]

Content

2. Let $\exists^*\zeta$ be $(\zeta \;!\; \exists)$ if ζ is a type and ζ if ζ is a proposition. The *α-content* of a mental state, ms, in symbols *alpha*-content(ms) is

$$\exists^*(ty_1\ f_1) \wedge \ldots \wedge \exists^*(ty_n\ f_n)$$

where $ms \models \langle\langle \alpha, a_1, ty_1, f_1, t_1; 1\rangle\rangle$

$\dots$

$ms \models \langle\langle \alpha, a_n, ty_n, f_n, t_n; 1\rangle\rangle$

and there is no other infon σ with relation α such that $ms \models \sigma$

Exportation

3. If

$$ms \models \alpha(a, \begin{array}{|l|}\hline \rho_1 \to X_1, \ \ldots, \ \rho_n \to X_n \\ \hline p(b) \\ \hline \end{array}, f, t)$$

then

$$\exists r(ms \models \alpha(a, \begin{array}{|l|}\hline \rho \to X, \rho_1 \to X_1, \ \ldots, \ \rho_n \to X_n \\ \hline p(X) \\ \hline \end{array},$$

$$f \cup [\rho \to b], t))$$

Different logics for different attitudes

4. $ms \models \langle\langle \alpha, a, ty, f, t;\ 1\rangle\rangle$ and
$ty \approx_\alpha ty'$
implies
$ms \models \langle\langle \alpha, a, ty', f, t;\ 1\rangle\rangle$
For example, if a believes that the glass is half full we would probably want it to follow that a believes that the glass is half empty.

[2] The need for distinguishing between an *internal* attitude predicate and the attitude predicates used in attitude reports is discussed in section 11.2.

However, if a is glad that the glass is half full it does not seem to follow that a is glad that the glass is half empty.

Rationality

5. If ms is rational then there is no infon $\sigma = \langle\!\langle \alpha, a, ty, f; 1\rangle\!\rangle$ such that

$$ms \models \sigma \text{ and } ty \text{ is } \begin{array}{|l||l|} \hline \multicolumn{2}{|c|}{\rho_1 \to X_1, \ldots, \rho_n \to X_n} \\ \hline \boxed{s} & \\ \sigma & \\ \overline{\sigma} & rest \\ \ldots & \\ \hline \end{array}, \quad \begin{array}{|l||l|} \hline \multicolumn{2}{|c|}{\rho_1 \to X_1, \ldots, \rho_n \to X_n} \\ \hline \boxed{\begin{array}{l} \boxed{s} \\ \sigma \end{array}} & \\ & rest \\ \boxed{\begin{array}{l} \boxed{s} \\ \overline{\sigma} \end{array}} & \\ \ldots & \\ \hline \end{array},$$

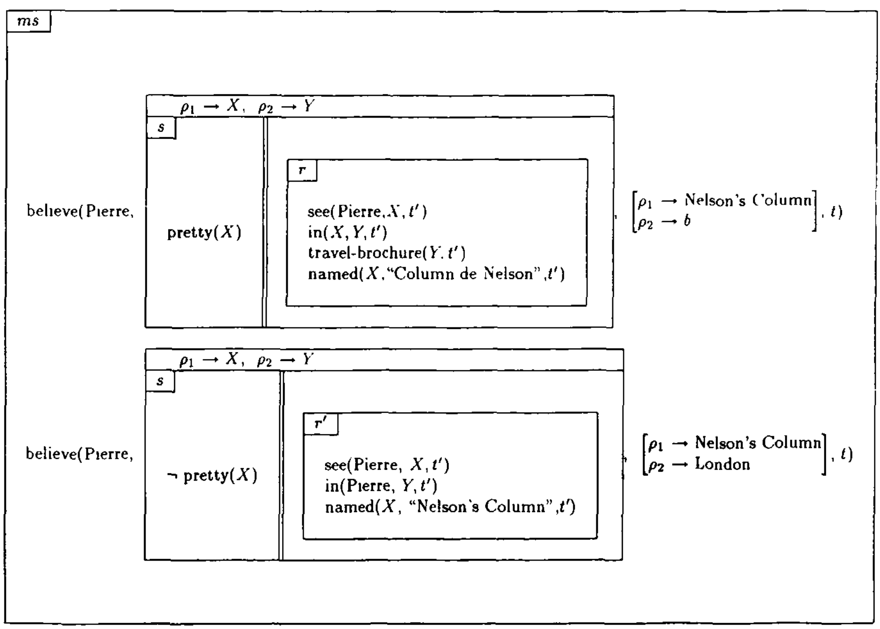
$$\begin{array}{|l||l|} \hline \multicolumn{2}{|c|}{\rho_1 \to X_1, \ldots, \rho_n \to X_n} \\ \hline p & \\ \neg\, p & rest \\ \ldots & \\ \hline \end{array}, \quad \begin{array}{|l||l|} \hline \multicolumn{2}{|c|}{\rho_1 \to X_1, \ldots, \rho_n \to X_n} \\ \hline X_i \neq X_i & rest \\ \ldots & \\ \hline \end{array}, \ldots.$$

11.1.2 Tackling the Puzzles

Kripke 1979 contains the by now well worn puzzle about Pierre, who believes that London is ugly but Londres is beautiful. Kripke argues that on the basis of his behaviour as a French speaker we appear to conclude that Pierre believes the proposition that London is beautiful, while on the basis of his behaviour as an English speaker we conclude that he does not believe the proposition that London is beautiful. Kripke's Pierre has spawned a cottage industry of related puzzles (see below). If we make the distinction between internal and external aspects of the belief, this gives us a finer grain than just propositions. There can be several different ty and f such that the result of applying ty to f all yield the same proposition. Although various diagnoses have been made concerning such puzzles, we believe an important insight in this regard is Donnellan 1990, who argues that '[...] the puzzles are about what it is possible for someone to believe or disbelieve in a situation and not upon Kripke's principles about sentences which *express* beliefs or upon a principle about translation.' (Donnellan 1990 p. 209). If this is correct, the current proposal is well placed to distinguish beliefs, since the technique employed, based on distinguishing between the restrictions of two given mental state types, is quite general and not restricted to linguistically based differentiation. (see e.g. Richard 1990).

We illustrate this by showing how to treat the Pierre puzzle in terms of mental states (using the situation theory and graphical notation developed in Barwise and Cooper 1991, Barwise and Cooper 1993).[3]

Pierre



[3]Actually, this example is for a version of Pierre that concerns the issue of whether Nelson's Column is pretty or not, a version that is more plausibly treated as involving contradictory beliefs about a single situation than Kripke's original formulation.

11.2 A Compositional Treatment

11.2.1 Bringing Propositions Back in

We have, then, two important tools for semantic analysis: propositions and mental states. However, as far as we are aware, there have been few if any attempts at incorporating mental states in a compositional semantics for a fragment that includes attitude reports. Indeed, Crimmins and Perry have expressed pessimism about the viability of such an attempt: 'Also closed is the prospect of a strictly compositional semantics for belief sentences. The semantic values of the subexpressions in a belief report, on our analysis, do not provide all the materials for the semantic value of the report itself.' (Crimmins and Perry 1989 p. 24).

Here we attempt to dispell such pessimism, though, in fairness to Crimmins and Perry, the view of compositionality they appear to be taking is compositionality of *content* rather than *meaning*. In so doing, however, we argue for a move that veers back at least part way to Montague's original analysis (Montague 1974): in common with Montague, we analyze 'believe' and its ilk as a relation that involves *at the very least* an agent, a time, and a proposition. There exist many basic arguments that demonstrate that *propositions have their place*, one which was denied them in the Barwise and Perry 1983 analysis of 'believe' as a relation between an agent, a time, and a (structured) mental state.[4]

We review briefly some arguments for an analysis of belief in terms of propositions:

(1a) Jill believes that p; her belief is true/false.

(1b) Jill believes Mike's claim/theory; Hence, Jill believes that Mike's claim/theory is true.

(1c) Jill believes everything Mike says. Mike says that Bill is here. Hence, Jill believes that Bill was here.

(1a) illustrates that beliefs are entities of which truth/falsity can be predicated. (1b) is one illustration of a more general phenomenon: all NP complements of 'believe' denote entities of which truth/falsity can be predicated; predicating that such an entity is believed involves predicating that such an entity is believed to be true. The simplest explanation of what's going on in (1c) (though certainly not the only one available cf. Forbes 1992.) is that whatever the cognitive argument of 'believe' is, it is identical to the complement of 'say' or 'assert', a highly plausible candidate for which is a proposition.

[4]But Barwise and Perry 1985 do seem to recognize the need: see p. 64.

11.2.2 An Option Substituting Austinian Propositions for Montague's Propositions

To what extent should one actually part from a Montague style analysis? At the present time we believe this question is to some extent an open one. Certainly given their independent motivation it seems quite natural and necessary to substitute *Austinian* propositions for Montague style propositions, even when the latter are recreated within situation theory (see Cooper 1993). Is this sufficient?

One might hope so, afterall whereas on most "neo-Russellian" accounts of propositions, there is just one proposition *that London is beautiful*, a situation semanticist armed with Austinian propositions can appeal to the existence of many such propositions, potentially as many such propositions as there are situations.

Unfortunately, such a hope seems to be frustrated because the situation in an Austinian proposition is the one which the proposition is about rather than the source or environment situation which represents something about how the agent comes to believe the proposition. Consider a case where Robin sees Anna go to school in the morning and believes later in the morning of a situation s at school that Anna is at school. However, later morning he is walking into work and sees a girl in the distance leaving the school. He does not realize that this girl is Anna but believes that the same situation s that supports Anna being at school supports the fact that this girl he sees is not at school. Clearly he believes that Anna is both at school and not at school. There is, however, no reason to require that he believes this of different situations.

The moral of this tale, it seems, is that Austinian propositions are not fine grained enough for the purposes at hand, namely distinguishing beliefs individuated by a rational agent. So we need to go a step further than merely replacing Montagovian propositions with Austinian ones.

The moral would appear to involve recognizing that "the mental" also has its place. This place can be located, we shall suggest, in at least two distinct ways: either by recognizing that attitudes possess an additional (implicit) argument, filled by or quantifying over one of the reported agent's mental states, a situation that, intuitively, reflects the currently reported perspective; or, more radically, for better and worse, by enriching the theory of truth for propositions along lines hinted in Barwise 1989 when he talks of Holmesian proposition. Both accounts interface onto the theory of mental states via constraints which relate beliefs in propositions to mental states.

We concentrate on developing the first of these two options.

11.3 Mental States as an Unarticulated Constituent

Our analysis is rooted in a situation semantic treatment of Montague's PTQ fragment.[5] We take the semantics of belief reports to express a relation between an agent a proposition and a mental state, i.e. a situation of the kind discussed in section 11.1. This is illustrated by our lexical entry for *believe.*

$[\underline{\alpha}]$ where α is a V_S

C

desc-sit $\rightarrow S$, desc-time $\rightarrow T$, $\langle$mental-state, $\underline{\alpha}\rangle \rightarrow MS$

$Prpn$

X

S

$\alpha'(X, Prpn, MS, T)$

C

The use of a propositional argument enables us to build up meanings for utterances in a standard compositional way since one can compositionally compute the meaning and potential content of the complement to *believe.* Constraints need to be placed on the relationship between the mental state (here provided by the context) and the proposition whose belief is reported. The most obvious and conservative ones for positive and negative attributions are expressed by the following constraints used in our fragment.

Attitude verbs

$$s \models \langle\langle BELIEVE, a, p, ms, t; 1\rangle\rangle \rightarrow$$
$$\exists T, f(ms \models \langle\langle BELIEVE\#, a, T, f, t; 1\rangle\rangle \wedge \exists^* Tf = p)$$

$$s \models \langle\langle BELIEVE, a, p, ms, t; 0\rangle\rangle \rightarrow$$
$$\neg\exists T, f(ms \models \langle\langle BELIEVE\#, a, T, f, t; 1\rangle\rangle \wedge \exists^* Tf = p)$$

(If α is a type $\exists^*\alpha$ is $(\alpha\ !\ \exists)$. If α is a proposition $\exists^*\alpha$ is α.)

The first constraint amounts to linking a positive belief attribution of proposition p relative to the mental situation ms with the existence of an *internal* belief state, classified by the relation BELIEVE#, such that applying its type component T and assignment component f yields p. The second constraint supplies the required analogue for negative belief attributions.

[5]For more details of this treatment see Cooper 1993.

We would like to suggest (in line with Barwise and Perry 1983) that the semantics only makes this requirement and that other considerations come into play in the pathological cases which require the meaning of the embedded sentence to be closer to the internal characterization of the mental state. An advantage of our theory of mental states is that the objects which we use to classify the internal aspects of mental states can be essentially similar to the objects which are used to characterize the meanings of sentences. This then enables such a theory to be made precise.

11.4 A Fragment with the Attitudes

Notation

Parameter Sorts

Parameters	Sort
S, MS	situation (mental state)
T	time
X, Y, X_i	individual
C, F	assignments (circumstances of an utterance)
$\mathcal{P}$	([X] $\rightarrow$ proposition) $\rightarrow$ proposition) *i.e.* type of types of individuals, a noun-phrase "content"
M	[C] $\rightarrow$ proposition *i.e.* a type of assignments (circumstances), a meaning
Prpn	proposition *i.e.* a sentence content
P, Q	[X] $\rightarrow$ proposition *i.e.* type of individuals

Combination ("Linguistic Application")

$\alpha\{\beta\} = \lambda[C](\alpha[C][\beta[C]])$

Uses

If α is a linguistic expression we use $\underline{\alpha}$ to represent a use or utterance of α.

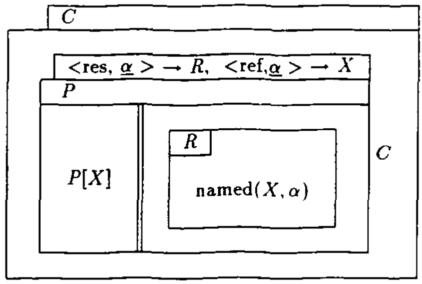
Assignments

$C/[\ r_1 \rightarrow a_1, \ldots, r_n \rightarrow a_n]$ is an assignment like C except it assigns a_i to r_i, for all i between 1 and n.

Lexicon

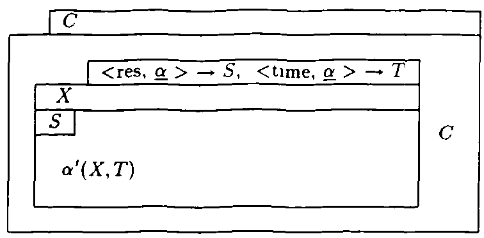
1 *proper names*

$[\![\alpha]\!]$ where α is a proper name



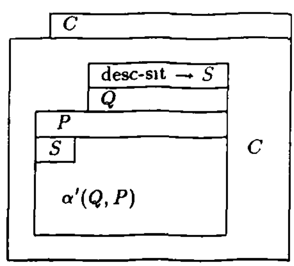
2 *common nouns*

$[\![\alpha]\!]$ where α is a common noun



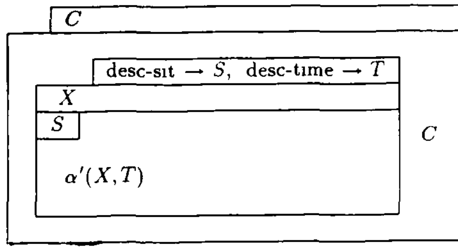
3 *determiners*

$[\![\alpha]\!]$ where α is a determiner



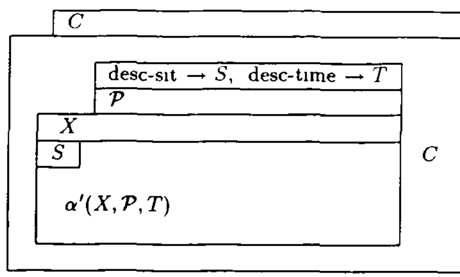
4. *intransitive verbs*

$[\![\alpha]\!]$ where α is an intransitive verb

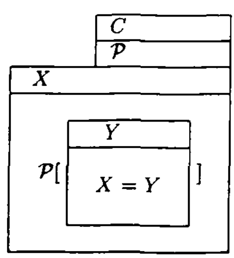


5. *transitive verbs*

$[\![\alpha]\!]$ where α is an transitive verb



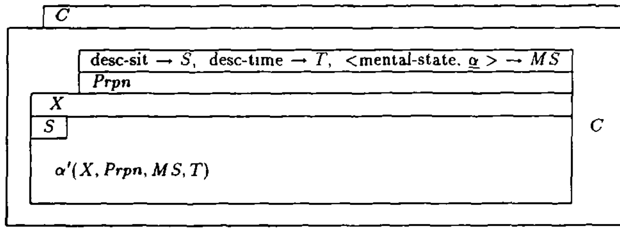
$[\![\text{be}]\!]$



Where: $(X = Y)$ abbreviates $(X, Y\ !\ =)$

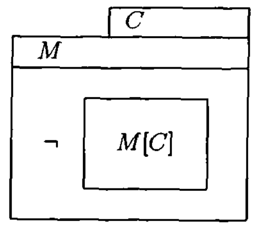
6. *verbs taking sentential complements*

$[\![\underline{\alpha}]\!]$ where α is a V_S

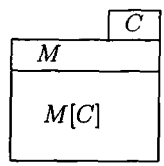


7. *Items used in Infl*

$[\![\underline{\text{n't}}]\!]$



$[\![\underline{\text{Pres}}]\!]$



8. *Variables (used in quantifying in constructions)*
   $[\![\underline{x_i}]\!]$ where x_i is a variable $\qquad X_i$

Phrase Structure

Non-branching rules are assumed to yield identical interpretations for their mother constituents as for their daughters unless explicitly mentioned otherwise.

1. *simple sentences*
   $[\![\underline{[_S\ \text{NP Infl}'\ \text{VP}]}]\!]$ $\qquad \lambda[C]([\![\underline{\text{Infl}'}]\!][C][[\![\underline{\text{NP}}]\!]\{[\![\underline{\text{VP}}]\!]\}])$

2. *noun phrases with determiner*
   $[\![\underline{[_{NP}\ \text{Det N}']}]\!]$ $\qquad [\![\underline{\text{Det}}]\!]\{[\![\underline{\text{N}'}]\!]\}$

3. *verb phrases with complements*
   $[\![\underline{[_{VP}\ V_t\ \text{NP}]}]\!]$ $\qquad [\![\underline{V_t}]\!]\{[\![\underline{\text{NP}}]\!]\}$

   $[\![\underline{[_{VP}\ V_S\ S']}]\!]$ $\qquad [\![\underline{V_S}]\!]\{[\![\underline{S'}]\!]\}$

4. *S′*

⟦[$_{S'}$ (that) Sent]⟧

C	
<concern, Sent> → *S* ⟦Sent⟧[*C*/[desc-sit → *S*]]	*C*

or

C	
∃*S*	⟦Sent⟧[*C*/[desc-sit → *S*]]

5. *Infl*

⟦[$_{Infl}$ Tns]⟧ ⟦Tns⟧

⟦[$_{Infl'}$ Infl n't]⟧

M / *C*
⟦n't⟧[*C*][⟦Infl⟧{[*M*]}]

6. *rules for quantifying in*

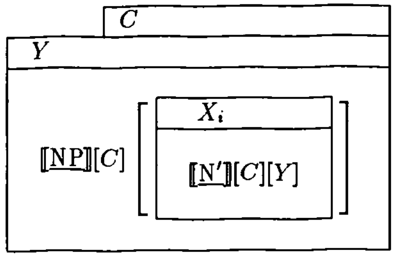
⟦[$_{S}$ NP X_i Sent]⟧

C		
⟦NP⟧[*C*][	X_i ⟦Sent⟧[*C*]	]

⟦[$_{VP}$ NP X_i VP]⟧

Y / *C*		
⟦NP⟧[*C*][	X_i ⟦VP⟧[C][Y]	]

$[\![{}_{\mathrm{N}'} \text{ NP } X_i \text{ N}']\!]$



Constraints

Determiners

1. $\exists s(s \models \langle\!\langle \text{every}, \tau_1, \tau_2; 1 \rangle\!\rangle)$ iff $\forall x \; x : \tau_1 \rightarrow x : \tau_2$
2. $\exists s(s \models \langle\!\langle \text{a}, \tau_1, \tau_2; 1 \rangle\!\rangle)$ iff $\exists x \; x : \tau_1 \wedge x : \tau_2$
3. $\exists s(s \models \langle\!\langle \text{the}, \tau_1, \tau_2; 1 \rangle\!\rangle)$ iff $\exists x(x : \tau_1 \wedge \forall y(y : \tau_1 \rightarrow y = x)) \wedge \forall x(x : \tau_1 \rightarrow x : \tau_2)$

We assume that there are only actual situations to give the $\Leftarrow$ direction of these biconditionals more that just modal force.

Extensional verbs

4. $s \models \langle\!\langle \alpha', x, \mathcal{P}, t; 1 \rangle\!\rangle$ iff
$$\mathcal{P}[\lambda[Y](s \; ! \; \langle\!\langle \alpha', x, Y, t; 1 \rangle\!\rangle)] \text{ is true.}$$

Attitude verbs

5. $s \models \langle\!\langle \alpha', a, p, ms, t; 1 \rangle\!\rangle \rightarrow$
$$\exists T, f(ms \models \langle\!\langle \alpha'\#, a, T, f, t; 1 \rangle\!\rangle \wedge \exists^* Tf = p$$

$s \models \langle\!\langle \alpha', a, p, ms, t; - \rangle\!\rangle \rightarrow$
$$\neg\exists T, f(ms \models \langle\!\langle \alpha'\#, a, T, f, t; + \rangle\!\rangle \wedge \exists^* Tf = p)$$

(If α is a type $\exists^*\alpha$ is $(\alpha \; ! \; \exists)$. If α is a proposition $\exists^*\alpha$ is α.)

References

Barwise, Jon. 1989. *The Situation in Logic.* CSLI Lecture Notes. Stanford: CSLI Publications.

Barwise, Jon, and Robin Cooper. 1991. Simple Situation Theory and its graphical representation. In *Partial and Dynamic Semantics III*, ed. Jerry Seligman. DYANA Deliverable R2.1.C. University of Edinburgh: Cemtre for Cognitive Science.

Barwise, Jon, and Robin Cooper. 1993. Extended Kamp Notation. In *Situation Theory and Its Applications, III*, ed. Peter Aczel, David Israel, Yasuhiro Katagiri, and Stanley Peters. CSLI Lecture Notes Number 37. Stanford: CSLI Publications.

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes.* Bradford Books. Cambridge: MIT Press.

Barwise, Jon, and John Perry. 1985. Shifting Situations and Shaken Attitudes. *Linguistics and Philosophy* 8:p. 399–452.

Cooper, Robin. 1993. Towards a general semantic framework. *DYANA-2 Deliverable* R2.1.A.

Crimmins, Mark, and John Perry. 1989. The Prince and the Phone Booth: Reporting Puzzling Beliefs. *CSLI Report* 88-128.

Donnellan, Keith. 1990. Belief and the Identity of Reference. In *The Role of Content in Logic, Language and Thought*, ed. C.A. Anderson and J. Owens. CSLI Lecture Notes. Stanford: CSLI Publications.

Forbes, Graeme. 1992. What Speech Reports Mean. *Tulane University Ms* xx.

Kripke, Saul. 1979. A puzzle about Belief. In *Meaning and Use*, ed. Avishai Margalit. Synthese language Library. Dordrecht: Reidel.

Montague, Richard. 1974. The Proper treatment of Quantification in Ordinary English. In *Formal Philosophy*, ed. Richmond Thomason. New Haven: Yale UP.

Richard, Mark. 1990. *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them.* Bradford Books. Cambridge: MIT Press.

12

Intensional Verbs Without Type-Raising or Lexical Ambiguity

MARY DALRYMPLE, JOHN LAMPING, FERNANDO PEREIRA AND VIJAY SARASWAT

Introduction

We present an analysis of the semantic interpretation of intensional verbs such as *seek* that allows them to take direct objects of either individual or quantifier type, producing both *de dicto* and *de re* readings in the quantifier case, all without needing to stipulate type-raising or quantifying-in rules. This simple account follows directly from our use of logical deduction in linear logic to express the relationship between syntactic structures and meanings. While our analysis resembles current categorial approaches in important ways (Moortgat 1988, Moortgat 1992a, Morrill 1993, Carpenter 1993), it differs from them in allowing the greater type flexibility of categorial semantics (van Benthem 1991) while maintaining a precise connection to syntax. As a result, we are able to provide derivations for certain readings of sentences with intensional verbs and complex direct objects that are not derivable in current purely categorial accounts of the syntax-semantics interface. The analysis forms a part of our ongoing work on semantic interpretation within the framework of Lexical-Functional Grammar.

12.1 Theoretical Background

As a preliminary to presenting our analysis of intensional verbs, we outline our approach to semantic interpretation in LFG.

It is well known that surface constituent structure does not always pro-

---

We thank David Israel, Michael Moortgat and Stanley Peters for discussions on the subject of this paper.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

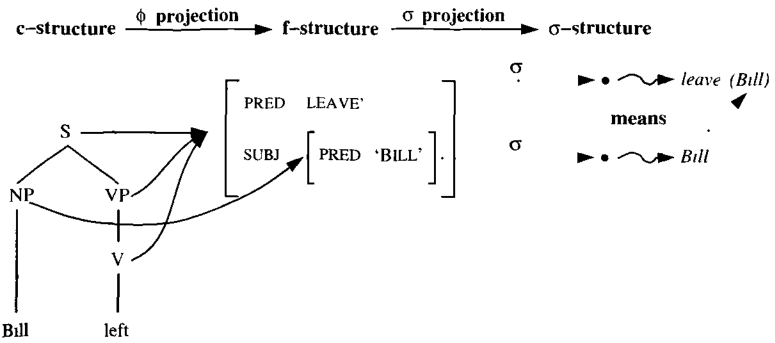


FIGURE 1 Semantic Interpretation Architecture

vide the optimal set of constituents or hierarchical structure to guide semantic interpretation. This has led to efforts to develop more abstract structures for the representation of relevant syntactic information. We follow Kaplan and Bresnan (1982) and Halvorsen:LI in taking the functional structure or *f-structure* of LFG as the primary input to semantic interpretation. The syntactic structures of LFG, the constituent structure or *c-structure* and the f-structure, are related by means of a functional correspondence, represented in Figure 1 by solid lines leading from nodes of the c-structure tree to f-structures (Kaplan and Bresnan 1982).[1] In more recent work, Kaplan:3Sed and HalvorsenKaplan:Projections have proposed to extend the theory of correspondences to other structures, called *projections*. Here, we will appeal to a semantic projection σ, relating f-structures and their meanings. Notationally, a subscript σ will indicate the semantic or σ projection of an f-structure f, so that the semantic projection of f will be written f_σ. In Figure 1, dotted lines represent the relation between f-structures and their semantic projections. Finally, as shown in the figure, the semantic projection f_σ of an f-structure f can be put in correspondence with a meaning ϕ:

(1) $f_\sigma \rightsquigarrow \phi$

Informally, we read this expression as "the meaning of f is ϕ". We use expressions of this sort to lexically associate meanings with f-structures, as in the following lexical entry for the word *Bill*:

(2) Bill $(\uparrow \text{PRED}) = \text{'BILL'}$
$\uparrow_\sigma \rightsquigarrow \textit{Bill}$

[1]The c-structure and f-structure presented here have been simplified to show only the detail necessary for the semantic issues addressed here. We also do not address a number of orthogonal semantic issues (tense and aspect, for example), providing only enough details of the representation of the meaning of a sentence to illustrate the points relevant to the discussion at hand.

The first line of this lexical entry:

$$(\uparrow \text{PRED}) = \text{`BILL'}$$

says (roughly) that the word *Bill* introduces an f-structure $\uparrow$ whose PRED is 'Bill'. The second line:

$$\uparrow_\sigma \rightsquigarrow Bill$$

says that the meaning of that f-structure is *Bill*. When a lexical entry is used, the metavariable '$\uparrow$' is instantiated with some constant f denoting an f-structure (Kaplan and Bresnan 1982, page 183). In particular, the term $\uparrow_\sigma$ is instantiated to the semantic projection f_σ of f and the formula $\uparrow_\sigma \rightsquigarrow Bill$ is instantiated to $f_\sigma \rightsquigarrow Bill$, asserting that the meaning of f is *Bill*.

More complicated lexical entries give not only meanings for f-structures, but also instructions for assembling f-structure meanings from the meanings of other f-structures. We distinguish a *meaning language*, in which we represent the meanings of f-structures, and a *composition language* or *glue language*, in which we describe how to assemble the meanings of f-structures from the meanings of their substructures. Each lexical entry will contain a composition language formula, its *meaning constructor*, specifying how a lexical entry contributes to the meaning of any structure in which it participates.

In principle, the meaning language can be any suitable logic. Here, since we are concerned with the semantics of intensional verbs, we will use Montague's higher-order intensional logic IL. The expressions that appear on the right side of the $\rightsquigarrow$ connective in lexical entries like (2) above are (usually open) terms in the meaning language.

Our composition language is a fragment of linear logic with higher-order quantification. While the resource sensitivity of linear logic is crucial to our overall interpretation framework, it does not play a central role in the analysis discussed here, so the linear connectives can be informally read as their classical counterparts.[2] In contrast to standard approaches, which use the λ-calculus to combine fragments of meaning via ordered applications, we combine fragments of meaning through unordered conjunction and impli-

[2]We make use of *linear logic* (Girard 1987) because its resource sensitivity turns out to be a good match to natural language This property of linear logic nicely captures, among other things, the LFG requirements of *coherence* and *consistency*, and enables a nice treatment of modification (Dalrymple et al 1993), quantification (Dalrymple et al 1994), and complex predicates (Dalrymple et al 1993) We make use only of the *tensor fragment* of linear logic The fragment is closed under conjunction, universal quantification and implication It arises from transferring to linear logic the ideas underlying the concurrent constraint programming scheme of Saraswat (1989) A nice tutorial introduction to linear logic itself may be found in Scedrov (1993)

cation. Rather than using λ-reduction to simplify meanings, we rely on deduction, as advocated by Pereira (1990, 1991).

12.2 A Simple Example

We now turn to a simple example to illustrate the framework. The lexical entries necessary for the example in Figure 1 are:[3]

(3) Bill $(\uparrow \text{PRED}) = \text{'BILL'}$
$\uparrow_\sigma = \mathit{Bill}$

left $(\uparrow \text{PRED}) = \text{'LEAVE'}$
$\forall X.\ (\uparrow \text{SUBJ})_\sigma \rightsquigarrow X \multimap \uparrow_\sigma \rightsquigarrow \mathit{leave}(X)$

The symbol '$\multimap$' is the *linear implication* operator of linear logic; for this paper, '$\multimap$' can be thought of as analogous to classical implication '$\rightarrow$'. The formula

$$\forall X.\ (\uparrow \text{SUBJ})_\sigma \rightsquigarrow X \multimap \uparrow_\sigma \rightsquigarrow \mathit{leave}(X)$$

states that the verb *left* requires a meaning X for its subject, $(\uparrow \text{SUBJ})$; when that meaning is provided, the meaning for the sentence will be $\mathit{leave}(X)$. When the words *Bill* and *left* are used in a sentence, the metavariable $\uparrow$ will be instantiated to a particular f-structure, and the meaning given in the lexical entry will be used as the meaning of that f-structure.

Here we repeat the f-structure in Figure 1, including labels for ease of reference:

(4)
$$f{:}\begin{bmatrix} \text{PRED} & \text{'LEAVE'} \\ \text{SUBJ} & g{:}\begin{bmatrix} \text{PRED} & \text{'BILL'} \end{bmatrix} \end{bmatrix}$$

Annotated phrase structure rules provide instructions for assembling this f-structure by instantiating the metavariables '$\uparrow$' in the lexical entries above. For instance, the metavariable '$\uparrow$' in the lexical entry for *Bill* is instantiated to the f-structure labeled g.

From the instantiated lexical entries of *Bill* and *left*, we have the following semantic information:

(5) **leave**: $\forall X.\ g_\sigma \rightsquigarrow X \multimap f_\sigma \rightsquigarrow \mathit{leave}(X)$
Bill: $g_\sigma \rightsquigarrow \mathit{Bill}$

where **leave** and **Bill** are names for their respective formulas. By modus ponens, we deduce:

$$\textbf{Bill}, \textbf{leave} \vdash f_\sigma \rightsquigarrow \mathit{leave}(\mathit{Bill})$$

[3] In the composition language, we use upper-case letters for *essentially existential* variables, that is, Prolog-like variables that become instantiated to particular terms during a derivation, and lower-case letters for *essentially universal variables* that stand for new local constants (also called eigenvariables) during a derivation.

The elements of the f-structure provide a set of formulas in the composition logic that introduce semantic elements and describe how they can be combined. For example, lexical items for words that expect arguments, like verbs, typically contribute a formula for combining the meanings of their arguments into a result. Once all the formulas are assembled, deduction in the logic is used to infer the meaning of the entire structure. Throughout this process we maintain a clear distinction between meanings proper and assertions about meaning combinations.

12.3 Quantification

We now turn to an overview of our analysis of quantification (Dalrymple, Lamping, Pereira, and Saraswat 1993). As a simple example, consider the sentence

(6) Every man left.

For conciseness, we will not illustrate the combination of the meaning constructors for *every* and *man*; instead, we will work with the derived meaning constructor for the subject *every man*, showing how it combines with the meaning constructor for *left* to produce a meaning constructor giving the meaning of the whole sentence.

The basic idea of our analysis of quantified NPs can be seen as a logical counterpart at the semantic composition level of the generalized-quantifier type assignment for (quantified) NPs (Barwise and Cooper 1981). Under that assignment, a NP meaning Q has type

$$(e \to t) \to t$$

—that is, a function from a property, the scope of quantification, to a proposition. At the semantic composition level, we can understand that type as follows. If by assuming that x is the entity referred to by the NP we can derive Sx as the meaning of the scope of quantification, where S is a property (a function from entities to propositions), then we can derive QS as the meaning of the whole clause containing the NP.

The f-structure for the sentence *Every man left* is:

(7)
$$f{:}\begin{bmatrix} \text{PRED} & \text{`LEAVE'} \\ \text{SUBJ} & g{:}\begin{bmatrix} \text{SPEC} & \text{`EVERY'} \\ \text{PRED} & \text{`MAN'} \end{bmatrix} \end{bmatrix}$$

The meaning constructors for *every man* and *left* are:[4]

[4]We use throughout the convenient abbreviation $Q(x, Rx, Sx)$ for the application of the generalized quantifier Q to restriction R and scope S.

(8) **leave**: $\forall X.\ g_\sigma \rightsquigarrow X \multimap f_\sigma \rightsquigarrow \mathit{leave}(X)$

every-man: $\forall H, S.\ (\forall x. g_\sigma \rightsquigarrow x \multimap H \rightsquigarrow Sx) \multimap H \rightsquigarrow \mathit{every}(z, \mathit{man}(z), Sz)$

The meaning constructor for *left* is as before. The meaning constructor for *every man* quantifies over semantic projections H which constitute possible quantification scopes; its propositional structure corresponds to the standard type assignment, $(e \to t) \to t$. It can be paraphrased as:

$\forall H, S.$

$(\forall x.\ g_\sigma \rightsquigarrow x$	if, by assuming an arbitrary meaning x for g,
$\multimap H \rightsquigarrow Sx)$	a meaning Sx for some scope H can be derived,
$\multimap H \rightsquigarrow \mathit{every}(z, \mathit{man}(z), Sz)$	then a possible complete meaning for H can be derived.

In the case at hand, the semantic projection f_σ will be chosen to provide the scope of quantification.[5] It has exactly the form that the antecedent of **every-man** expects. The meaning S will be instantiated to $\lambda x.\mathit{leave}(x)$. From the premises in (8), we can deduce the meaning of the scope f-structure f:

$$\textbf{every-man}, \textbf{leave} \vdash f_\sigma \rightsquigarrow \mathit{every}(z, \mathit{man}(z), \mathit{leave}(z))$$

The resource sensitivity of linear logic ensures that the scope of quantification is constructed and used exactly once.

12.4 Intensional Verbs

We follow Montague (1974) in requiring intensional verbs like *seek* to take an object of NP type. What is interesting is that this is the only step required in our setting to obtain the appropriate ambiguity predictions for intensional verbs. The *de re*/*de dicto* ambiguity of a sentence like *Bill seeks a unicorn*:

de dicto reading: $\mathit{seek}(\mathit{Bill}, {}^\wedge\lambda Q.a(x, \mathit{unicorn}(x), [{}^\vee Q](x)))$
de re reading: $a(x, \mathit{unicorn}(x), \mathit{seek}(\mathit{Bill}, {}^\wedge\lambda Q.[{}^\vee Q](x)))$

is a natural consequence, in our setting, of *seek* taking an NP-type argument.

[5] From what has been said so far, g_σ could also be chosen to provide the scope, leading to a nonsensical result. As explained in our full analysis of quantifiers (Dalrymple, Lamping, Pereira, and Saraswat 1993), that problem is avoided by using a family of $\rightsquigarrow$ relations indexed by the type of their second argument. The relation for the meaning of the scope of quantification is the one that expects a proposition meaning, so g_σ can not provide a scope.

We assign the following lexical entry to the verb *seek*:

(9)
$$
\begin{array}{l}
\text{seek} \;\; (\uparrow \textsc{pred}) = \text{`SEEK'} \\
\quad \forall Z, Y.\;\; (\uparrow \textsc{subj})_\sigma \rightsquigarrow Z \\
\qquad\quad \otimes (\forall s, p.(\forall X.(\uparrow \textsc{obj})_\sigma \rightsquigarrow X \multimap s \rightsquigarrow p(X)) \multimap s \rightsquigarrow Y({}^\wedge p)) \\
\qquad \multimap\; \uparrow_\sigma \rightsquigarrow \mathit{seek}(Z, {}^\wedge Y)
\end{array}
$$

The significant fact here is that *seek* differs from an extensional verb such as *find* below (corresponding to the type $e \to e \to t$) in its specification of requirements on its object:

(10)
$$
\begin{array}{l}
\text{find} \;\; (\uparrow \textsc{pred}) = \text{`FIND'} \\
\quad \forall Z, Y.(\uparrow \textsc{subj})_\sigma \rightsquigarrow Z \otimes (\uparrow \textsc{obj})_\sigma \rightsquigarrow Y \multimap\; \uparrow_\sigma \rightsquigarrow \mathit{find}(Z, Y)
\end{array}
$$

Note also the use of the operators "${}^\wedge$" and "${}^\vee$" of IL. Computationally, this implies that our proofs have to be carried out in a logic whose terms are (typed) lambda-expressions that satisfy $\alpha-$, $\beta-$ and η-equality and also the law ${}^\vee({}^\wedge P) = P$, for all P.

The lexical entry for *seek* can be paraphrased as follows:

$\forall Z, Y.\ (\uparrow \textsc{subj})_\sigma \rightsquigarrow Z \otimes$	$\left\{\right.$ The verb *seek* requires a meaning Z for its subject and	
$(\forall s, p.\ (\forall X.\ (\uparrow \textsc{obj})_\sigma \rightsquigarrow X \multimap s \rightsquigarrow p(X)) \multimap s \rightsquigarrow Y({}^\wedge p))$	$\left\{\right.$ a meaning ${}^\wedge Y$ for its object, where Y is an NP meaning applied to the meaning p of an arbitrarily-chosen 'scope' s,	$(*)$
$\multimap\; \uparrow_\sigma \rightsquigarrow \mathit{seek}(Z, {}^\wedge Y)$	$\left\{\right.$ to produce the clause meaning $\mathit{seek}(Z, {}^\wedge Y)$.	

Rather than looking for an entity type meaning for its object, the requirement expressed by the subformula labeled $(*)$ exactly describes the form of quantified NP meanings discussed in the previous section. In this case, a quantified NP in the object position is one that can accept anything that takes a meaning for $(\uparrow \textsc{obj})_\sigma$ to a meaning for any s, and convert that into a meaning for the s. In particular, the quantified NP *a unicorn* will fill the requirement, as we now demonstrate.

The f-structure for *Bill seeks a unicorn* is:

(11)
$$
f{:}\left[\begin{array}{lll}
\textsc{pred} & \text{`SEEK'} & \\
\textsc{subj} & g{:}\left[\begin{array}{ll}\textsc{pred} & \text{`BILL'}\end{array}\right] & \\
\textsc{obj} & h{:}\left[\begin{array}{ll}\textsc{spec} & \text{`A'} \\ \textsc{pred} & \text{`UNICORN'}\end{array}\right] &
\end{array}\right]
$$

The semantic information associated with this f-structure is:

$$\begin{array}{ll} \textbf{seeks}: & \forall Z, Y.\ \ g_\sigma \rightsquigarrow Z \\ & \quad\quad \otimes(\forall s, p.(\forall X.h_\sigma \rightsquigarrow X \multimap s \rightsquigarrow p(X)) \multimap s \rightsquigarrow Y(\hat{}p)) \\ & \quad \multimap f_\sigma \rightsquigarrow \mathit{seek}(z, \hat{}Y) \\ \textbf{Bill}: & g_\sigma \rightsquigarrow \mathit{Bill} \\ \textbf{a-unicorn}: & \forall H, S.(\forall x.h_\sigma \rightsquigarrow x \multimap H \rightsquigarrow Sx) \multimap H \rightsquigarrow a(z, \mathit{unicorn}(z), Sz) \end{array}$$

These are the premises for the deduction of the meaning of the sentence *Bill seeks a unicorn*. From the premises **Bill** and **seeks**, we can conclude by modus ponens:

$$\begin{array}{ll} \textbf{Bill-seeks}: & \forall Y.(\forall s, p.(\forall X.h_\sigma \rightsquigarrow X \multimap s \rightsquigarrow p(X)) \multimap s \rightsquigarrow Y(\hat{}p)) \\ & \quad\quad \multimap f_\sigma \rightsquigarrow \mathit{seek}(\mathit{Bill}, \hat{}Y) \end{array}$$

Different derivations starting from the premises **Bill-seeks** and **a-unicorn** will yield the different readings of *Bill seeks a unicorn* that we seek.

12.4.1 De Dicto Reading

The formula **a-unicorn** is exactly what is required by the antecedent of **Bill-seeks** provided that the following substitutions are performed:

$$\begin{array}{l} H \mapsto s \\ S \mapsto p \\ X \mapsto x \\ Y \mapsto \lambda P.a(z, \mathit{unicorn}(z), [\check{}P](z)) \end{array}$$

We can thus conclude the desired *de dicto* reading:

$$f_\sigma \rightsquigarrow \mathit{seek}(\mathit{Bill}, \hat{}\lambda P.a(z, \mathit{unicorn}(z), [\check{}P](z)))$$

To show how the premises also support a *de re* reading, we take first a short detour through a simpler example.

12.4.2 Nonquantified Objects

The meaning constructor for *seek* also allows for nonquantified objects as arguments, without needing a special type-raising rule. Consider the f-structure for the sentence *Bill seeks Al*:

(12) $$f{:}\left[\begin{array}{ll} \text{PRED} & \text{‘SEEK’} \\ \text{SUBJ} & g{:}\left[\begin{array}{ll}\text{PRED} & \text{‘BILL’}\end{array}\right] \\ \text{OBJ} & h\left[\begin{array}{ll}\text{PRED} & \text{‘AL’}\end{array}\right] \end{array}\right]$$

The lexical entry for *Al* is analogous to the one for *Bill*. We begin with the premises **Bill-seeks** and **Al**.

$$\begin{array}{ll} \textbf{Bill-seeks}: & \forall Y.(\forall s, p.(\forall X.h_\sigma \rightsquigarrow X \multimap s \rightsquigarrow p(X)) \multimap s \rightsquigarrow Y(\hat{}p)) \\ & \quad\quad \multimap f_\sigma \rightsquigarrow \mathit{seek}(\mathit{Bill}, \hat{}Y) \\ \textbf{Al}: & h_\sigma \rightsquigarrow \mathit{Al} \end{array}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{h_\sigma \rightsquigarrow Al \vdash h_\sigma \rightsquigarrow Al \qquad s \rightsquigarrow P(Al) \vdash s \rightsquigarrow P(Al)}{h_\sigma \rightsquigarrow Al, h_\sigma \rightsquigarrow Al \multimap s \rightsquigarrow P(Al) \vdash s \rightsquigarrow P(Al)}}{h_\sigma \rightsquigarrow Al, (\forall x. h_\sigma \rightsquigarrow x \multimap s \rightsquigarrow P(x)) \vdash s \rightsquigarrow P(Al)}}{h_\sigma \rightsquigarrow Al \vdash (\forall x. h_\sigma \rightsquigarrow x \multimap s \rightsquigarrow P(x)) \multimap s \rightsquigarrow P(Al)}}{h_\sigma \rightsquigarrow Al \vdash \forall P.(\forall x. h_\sigma \rightsquigarrow x \multimap s \rightsquigarrow P(x)) \multimap s \rightsquigarrow P(Al)}$$

FIGURE 2 Proof that **Al** can function as a quantifier

For the derivation to proceed, **Al** must supply the NP meaning constructor that **Bill-seeks** requires. This is possible because **Al** can map a proof Π of the meaning for s from the meaning for h into a meaning for s, simply by supplying $h_\sigma \rightsquigarrow Al$ to Π. Formally, from **Al** we can prove (Figure 2):

(13) $\forall P.(\forall x. h_\sigma \rightsquigarrow x \multimap s \rightsquigarrow P(x)) \multimap s \rightsquigarrow P(Al)$

This corresponds to the Montagovian type-raising of a proper name meaning to an NP meaning, and also to the undirected Lambek calculus derivation of the sequent $e \Rightarrow (e \rightarrow t) \rightarrow t$.

Formula (13) with the substitutions

$$P \mapsto p, Y \mapsto \lambda P.[{}^\vee P](Al)$$

can then be used to satisfy the antecedent of **Bill-seeks** to yield the desired result:

$$f_\sigma \rightsquigarrow \mathit{seek}(\mathit{Bill}, {}^\wedge\lambda P.[{}^\vee P](Al))$$

It is worth contrasting the foregoing derivation with treatments of the same issue in the lambda calculus. The function $\lambda x.\lambda P.Px$ raises a term like Al to the quantified NP form $\lambda P.P(Al)$, so it is easy to modify Al to make it suitable for **seek**. But the conversion must be explicitly applied somewhere, either in a meaning postulate or in an alternate definition for *seek*, in order to carry out the derivation. This is because a lambda calculus function must specify exactly what is to be done with its arguments and how they will interact. It must presume some functional form of its arguments in order to achieve its own function. Thus, it is impossible to write a function that is indifferent with respect to whether its argument is Al or $\lambda P.P(Al)$.

In the deductive framework, on the other hand, the exact way in which different propositions can interact is not fixed, although it is constrained by their (logical and quantificational) propositional structure. Thus $h_\sigma \rightsquigarrow Al$ can function as any logical consequence of itself, in particular as:

$$\forall S, P. \ (\forall x. \ h_\sigma \rightsquigarrow x \multimap S \rightsquigarrow P(x)) \multimap S \rightsquigarrow P(Al)$$

This flexibility, which is also found in syntactic-semantic analyses based on the Lambek calculus and its variants (Moortgat 1988, Moortgat 1992b, van Benthem 1991), seems to align well with some of the type flexibility in natural language.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{I \rightsquigarrow Z \vdash I \rightsquigarrow Z \qquad S \rightsquigarrow P(Z) \vdash S \rightsquigarrow P(Z)}{I \rightsquigarrow Z, I \rightsquigarrow Z \multimap S \rightsquigarrow P(Z) \vdash S \rightsquigarrow P(Z)}}{I \rightsquigarrow Z, (\forall x.\, I \rightsquigarrow x \multimap S \rightsquigarrow P(x)) \vdash S \rightsquigarrow P(Z)}}{I \rightsquigarrow Z \vdash (\forall x.\, I \rightsquigarrow x \multimap S \rightsquigarrow P(x)) \multimap S \rightsquigarrow P(Z)}}{I \rightsquigarrow Z \vdash \forall S, P.\, (\forall x.\, I \rightsquigarrow x \multimap S \rightsquigarrow P(x)) \multimap S \rightsquigarrow P(Z)}}{\vdash I \rightsquigarrow Z \multimap \forall S, P.\, (\forall x.\, I \rightsquigarrow x \multimap S \rightsquigarrow P(x)) \multimap S \rightsquigarrow P(Z)}}{\vdash \forall I, Z.\, I \rightsquigarrow Z \multimap \forall S, P.\, (\forall x.\, I \rightsquigarrow x \multimap S \rightsquigarrow P(x)) \multimap S \rightsquigarrow P(Z)}$$

FIGURE 3 General Type-Raising Theorem

12.4.3 Type Raising and Quantifying In

The derivation in Figure 2 can be generalized as shown in Figure 3 to produce the general type-raising theorem:

(14) $\forall I, Z.\, I \rightsquigarrow Z \multimap (\forall S, P.\, (\forall x. I \rightsquigarrow x \multimap S \rightsquigarrow P(x)) \multimap S \rightsquigarrow P(Z))$

This theorem can be used to raise meanings of e type to $(e \to t) \to t$ type, or, dually, to quantify into verb argument positions. For example, with the variable instantiations

$$\begin{aligned} I &\mapsto h_\sigma \\ X &\mapsto x \\ P &\mapsto p \\ S &\mapsto s \\ Y &\mapsto \lambda R.[{}^{\vee}R](Z) \end{aligned}$$

we can use transitivity of implication to combine (14) with **Bill-seeks** to derive:

$$\textbf{Bill-seeks}'\!: \forall Z.\, h_\sigma \rightsquigarrow Z \multimap f_\sigma \rightsquigarrow \mathit{seek}(\mathit{Bill}, {}^{\wedge}\lambda R.[{}^{\vee}R](Z))$$

This formula can then be combined with arguments of type e to produce a meaning for f_σ. For instance, it will take the non-type-raised $h_\sigma \rightsquigarrow Al$ to yield the same result

$$f_\sigma \rightsquigarrow \mathit{seek}(\mathit{Bill}, {}^{\wedge}\lambda R.[{}^{\vee}R](Al))$$

as the combination of **Bill-seeks** with the type-raised version of **Al**. In fact, **Bill-seeks**$'$ corresponds to type $e \to t$, and can thus be used as the scope of a quantifier, which would then quantify into the intensional direct object argument of *seek*. As we will presently see, that is exactly what is needed to derive *de re* readings.

12.4.4 De Re Reading

We have just seen how theorem (14) provides a general mechanism for quantifying into intensional argument positions. In particular, it allowed the derivation of **Bill-seeks′** from **Bill-seeks**. Now, given the premises

$$\begin{array}{ll}\textbf{Bill-seeks}': & \forall Z.\ h_\sigma \rightsquigarrow Z \multimap f_\sigma \rightsquigarrow \mathit{seek}(\mathit{Bill}, {}^\wedge\lambda R.[{}^\vee R](Z)) \\ \textbf{a-unicorn}: & \forall H, S.\ (\forall x.\ h_\sigma \rightsquigarrow x \multimap H \rightsquigarrow Sx) \multimap H \rightsquigarrow a(z, \mathit{unicorn}(z), Sz)\end{array}$$

and the variable substitutions

$$Z \mapsto x$$

$$H \mapsto f_\sigma$$

$$S \mapsto \lambda z.\mathit{seek}(\mathit{Bill}, {}^\wedge\lambda R.[{}^\vee R](z))$$

we can apply modus ponens to derive the *de re* reading of *Bill seeks a unicorn*:

$$f_\sigma \rightsquigarrow a(z, \mathit{unicorn}(z), \mathit{seek}(\mathit{Bill}, {}^\wedge\lambda R.[{}^\vee R](z)))$$

12.4.5 A Comparison with Categorial Approaches

The analysis presented here has strong similarities to analyses of the same phenomena discussed by Morrill:type-logical and Carpenter:quant-scope. Following Moortgat:discontinuous, they add to an appropriate version of the Lambek calculus (Lambek 1958) the *scope* connective $\Uparrow$, subject to the following proof rules:

$$\frac{\Gamma, v : A, \Gamma' \Rightarrow u : B \qquad \Delta, t(\lambda v.u) : B, \Delta' \Rightarrow C}{\Delta, \Gamma, t : A \Uparrow B, \Gamma', \Delta' \Rightarrow C}\ [\mathrm{QL}]$$

$$\frac{\Gamma \Rightarrow u : A}{\Gamma \Rightarrow \lambda v.v(u) : A \Uparrow B}\ [\mathrm{QR}]$$

In terms of the scope connective, a quantified noun phrase is given the category N $\Uparrow$ S, which semantically corresponds to the type $(e \to t) \to t$ and agrees with the propositional structure of our linear formulas for quantified noun phrases, for instance (8). A phrase of category $N \Uparrow S$ is an infix functor that binds a variable of type e, the type of individual noun phrases N, within a scope of type t, the type of sentences S. An intensional verb like 'seek' has, then, category (N \ (S)/(N $\Uparrow$ S), with corresponding type $(((e \to t) \to t) \to e \to t)$. [6] Thus the intensional verb will take as direct object a quantified noun phrase, as required.

A problem arises, however, with sentences such as

(15) Bill seeks a conversation with every unicorn.

This sentence has five possible interpretations:

(16) a. $\mathit{seek}(\mathit{Bill}, {}^\wedge\lambda P.\mathit{every}(u, \mathit{unicorn}(u), a(z, \mathit{conv\text{-}with}(z, u), [{}^\vee P](z))))$
b. $\mathit{seek}(\mathit{Bill}, {}^\wedge\lambda P.a(z, \mathit{every}(u, \mathit{unicorn}(u), \mathit{conv\text{-}with}(z, u)), [{}^\vee P](z)))$

[6]These category and type assignments are an oversimplification since intensional verbs like 'seek' require a direct object of type $s \to ((e \to t) \to t)$, but for the present discussion the simpler category and type are sufficient. Morrill:type-logical provides a full treatment.

c. $every(u, unicorn(u), seek(Bill, {}^\wedge\lambda P.a(z, conv\text{-}with(z, u), [{}^\vee P](z))))$
d. $every(u, unicorn(u), a(z, conv\text{-}with(z, u), seek(Bill, {}^\wedge\lambda P.[{}^\vee P](z))))$
e. $a(z, every(u, unicorn(u), conv\text{-}with(z, u)), seek(Bill, {}^\wedge\lambda P.[{}^\vee P](z)))$

Both our approach and the categorial analysis using the scope connective have no problem in deriving interpretations (16b), (16c), (16d) and (16e). In those cases, the scope of 'every unicorn' is interpreted as an appropriate term of type $e \to t$. However, the situation is different for interpretation (16a), in which both the conversations and the unicorn are *de dicto*, but the conversations sought may be different for different unicorns sought. As we will show below, this interpretation can be easily derived within our framework. However, a similar derivation does not appear possible in terms of the categorial scoping connective.

The difficulty for the categorial account is that the category N $\Uparrow$ S represents a phrase that plays the role of a category N phrase where it appears, but takes an S (dependent on the N) as its scope. In the derivation of (16a), however, the scope of 'every unicorn' is 'a conversation with', which is not of category S. Semantically, 'a conversation with' is represented by:

(17) $\lambda P.\lambda u.a(z, conv\text{-}with(z, u), [{}^\vee P](z)) : (e \to t) \to (e \to t)$

The *undirected* Lambek calculus (van Benthem 1991) allows us to compose (17) with the interpretation of 'every unicorn':

(18) $\lambda Q.every(u, unicorn(u), Q(u)) : (e \to t) \to t$

to yield:

(19) $\lambda P.every(u, unicorn(u), a(z, conv\text{-}with(z, u), [{}^\vee P](z))) : (e \to t) \to t$

As we will see below, our linear logic formulation also allows that derivation step.

In contrast, as Moortgat:discontinuous points out, the categorial rule [QR] is not powerful enough to raise $N \Uparrow S$ to take as scope any functor whose result is a S. In particular, the sequent

(20) $\mathrm{N} \Uparrow \mathrm{S} \Rightarrow \mathrm{N} \Uparrow (\mathrm{N} \Uparrow \mathrm{S})$

is not derivable, whereas the corresponding "semantic" sequent (up to permutation)

(21) $q : (e \to t) \to t \Rightarrow$
$\lambda R.\lambda P.q(\lambda x.R(P)(x)) : ((e \to t) \to (e \to t)) \to ((e \to t) \to t)$

is derivable in the undirected Lambek calculus. Sequent (21) will in particular raise (18) to a function that, applied to (17), produces (19), as required.

Furthermore, the solution proposed by Morrill:type-logical to make the scope calculus complete is to restrict the intended interpretation of $\Uparrow$ so that

(20) is not valid. Thus, *contra* Carpenter:quant-scope, Morrill's logically more satisfying account of $\Uparrow$ is not a step towards making reading (16a) available.

We now give the derivation of the interpretation (16a) in our framework. The f-structure for (15) is:

(22)
$$f:\left[\begin{array}{ll} \text{PRED} & \text{`SEEK'} \\ \text{SUBJ} & g:\left[\begin{array}{ll}\text{PRED} & \text{`BILL'}\end{array}\right] \\ \text{OBJ} & h:\left[\begin{array}{ll} \text{SPEC} & \text{`A'} \\ \text{PRED} & \text{`CONVERSATION'} \\ \text{OBL}_{\text{WITH}} & i:\left[\begin{array}{ll}\text{SPEC} & \text{`EVERY'} \\ \text{PRED} & \text{`UNICORN'}\end{array}\right] \end{array}\right] \end{array}\right]$$

The two formulas **Bill-seeks** and **every-unicorn** can be derived as described before:

Bill-seeks: $\forall Y.(\forall s, p.(\forall X.h_\sigma \rightsquigarrow X \multimap s \rightsquigarrow pX) \multimap s \rightsquigarrow Y(\hat{}p))$
$\multimap f_\sigma \rightsquigarrow \mathit{seek}(\mathit{Bill}, \hat{}Y)$

every-unicorn: $\forall G, S.(\forall x.i_\sigma \rightsquigarrow x \multimap G \rightsquigarrow Sx)$
$\multimap G \rightsquigarrow \mathit{every}(u, \mathit{unicorn}(u), Su)$

As explained in more detail in Dalrymple, Lamping, Pereira, and Saraswat (1993), the remaining lexical premises for (22) are:

a: $\forall H, R, T.((\forall x.(h_\sigma \text{VAR}) \rightsquigarrow x \multimap (h_\sigma \text{RESTR}) \rightsquigarrow Rx)$
$\otimes(\forall x.h_\sigma \rightsquigarrow x \multimap H \rightsquigarrow Tx))$
$\multimap H \rightsquigarrow a(z, Rz, Tz)$

conv-with: $\forall Z, X.\ (h_\sigma \text{VAR}) \rightsquigarrow Z \otimes i \rightsquigarrow X$
$\multimap (h_\sigma \text{RESTR}) \rightsquigarrow \mathit{conv\text{-}with}(Z, X)$

From these premises we immediately derive

$$\forall X, H, T.i_\sigma \rightsquigarrow X \otimes (\forall x.h_\sigma \rightsquigarrow x \multimap H \rightsquigarrow Tx))$$
$$\multimap H \rightsquigarrow a(z, \mathit{conv\text{-}with}(z, X), Tz)$$

which can be rewritten as:

(23) $\forall H, T.(\forall x.h_\sigma \rightsquigarrow x \multimap H \rightsquigarrow Tx) \multimap$
$\forall X.\ (i_\sigma \rightsquigarrow X \multimap H \rightsquigarrow a(z, \mathit{conv\text{-}with}(z, X), Tz))$

With the substitutions

$$X \mapsto x, G \mapsto H, S \mapsto \lambda v.a(z, \mathit{conv\text{-}with}(z, v), Tz))$$

formula (23) can be combined with **every-unicorn** to yield the required quantifier-type formula:

(24) $\forall H, T.(\forall x.h_\sigma \rightsquigarrow x \multimap H \rightsquigarrow Tx) \multimap$
$H \rightsquigarrow \mathit{every}(u, \mathit{unicorn}(u), a(z, \mathit{conv\text{-}with}(z, u), Tz))$

Using substitutions

$$\begin{array}{lll} H & \mapsto & s \\ T & \mapsto & p \\ Y & \mapsto & \lambda R.every(u, unicorn(u), a(z, conv\text{-}with(z, u), [{}^{\vee}R](z)))) \end{array}$$

and modus ponens, we then combine (24) with **Bill-seeks** to obtain the desired final result:

$$f_\sigma \rightsquigarrow seek(Bill, {}^{\wedge}\lambda R.every(u, unicorn(u), a(z, conv\text{-}with(z, u), [{}^{\vee}R](z))))$$

We see thus that our more flexible connection between syntax and semantics permits the full range of type flexibility provided categorial *semantics* without losing the rigorous connection to syntax. In contrast, current categorial accounts of the syntax-semantics interface do not appear to offer the needed flexibility when syntactic and semantic composition are more indirectly connected, as in the present case.

12.5 Conclusion

We have shown that our deductive framework allows us to predict the correct set of readings for intensional verbs with quantified and nonquantified direct objects if we make a single assumption: that intensional verbs require a quantified direct object. This assumption is, of course, the starting point of the standard Montagovian treatment of intensional verbs. But that treatment depends on the additional machinery of quantifying in to generate *de re* readings of quantified direct objects, and that of explicit type raising to accommodate unquantified direct objects. In our approach those problems are handled directly by the deductive apparatus without further stipulation.

These results, as well as our previous work on quantifier scope, suggest the advantages of a generalized form of compositionality in which the semantic contributions of phrases are represented by logical formulas rather than by functional abstractions as in traditional compositionality. The fixed application order and fixed type requirements of lambda terms are just too restrictive when it comes to encoding the freer order of information presentation in natural language. In this observation, our treatment is closely related to systems of syntactic and semantic type assignment based on the Lambek calculus and its variants. However, we differ from those categorial approaches in providing an explicit link between functional structures and semantic derivations that does not depend on linear order and constituency in syntax to keep track of predicate-argument relations. Thus we avoid the need to force both syntax and semantics into an uncomfortably tight categorial embrace.

References

Barwise, Jon, and Robin Cooper. 1981. Generalized Quantifiers and Natural Language. *Linguistics and Philosophy* 4:159–219.

Carpenter, Bob. 1993. Quantification and Scoping: a Deductive Account. MS, Carnegie-Mellon University.

Dalrymple, Mary, Angie Hinrichs, John Lamping, and Vijay Saraswat. 1993. The Resource Logic of Complex Predicate Interpretation. In *Proceedings of the 1993 Republic of China Computational Linguistics Conference (ROCLING)*. Hsitou National Park, Taiwan, September. Computational Linguistics Society of R.O.C.

Dalrymple, Mary, John Lamping, Fernando C. N. Pereira, and Vijay Saraswat. 1994. A Deductive Account of Quantification in LFG. In *Quantifiers, Deduction, and Context*, ed. Makoto Kanazawa, Christopher J. Piñón, and Henriette de Swart. Stanford, California: Center for the Study of Language and Information. To appear.

Dalrymple, Mary, John Lamping, and Vijay Saraswat. 1993. LFG Semantics Via Constraints. In *Proceedings of the Sixth Meeting of the European ACL*. University of Utrecht, April. European Chapter of the Association for Computational Linguistics.

Girard, J.-Y. 1987. Linear Logic. *Theoretical Computer Science* 45:1–102.

Halvorsen, Per-Kristian. 1983. Semantics for Lexical-Functional Grammar. *Linguistic Inquiry* 14(4):567–615.

Halvorsen, Per-Kristian, and Ronald M. Kaplan. 1988. Projections and Semantic Description in Lexical-Functional Grammar. In *Proceedings of the International Conference on Fifth Generation Computer Systems*, 1116–1122. Tokyo, Japan. Institute for New Generation Systems.

Kaplan, Ronald M. 1987. Three Seductions of Computational Psycholinguistics. In *Linguistic Theory and Computer Applications*, ed. Peter Whitelock, Harold Somers, Paul Bennett, Rod Johnson, and Mary McGee Wood. 149–188. London: Academic Press.

Kaplan, Ronald M., and Joan Bresnan. 1982. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In *The Mental Representation of Grammatical Relations*, ed. Joan Bresnan. 173–281. Cambridge, MA: The MIT Press.

Lambek, Joachim. 1958. The Mathematics of Sentence Structure. *American Mathematical Monthly* 65:154–170.

Montague, Richard. 1974. The Proper Treatment of Quantification in Ordinary English. In *Formal Philosophy*, ed. Richmond Thomason. New Haven: Yale University Press.

Moortgat, Michael. 1988. *Categorial Investigations: Logical and Linguistic Aspects of the Lambek Calculus*. Doctoral dissertation, University of Amsterdam, Amsterdam, The Netherlands.

Moortgat, Michael. 1992a. Generalized Quantifiers and Discontinuous Type Constructors. In *Discontinuous Constituency*, ed. W. Sijtsma and H. van

Horck. Berlin, Germany: Mouton de Gruyter. To appear.

Moortgat, Michael. 1992b. Labelled Deductive Systems for Categorial Theorem Proving. In *Proceedings of the Eighth Amsterdam Colloquium*, ed. P. Dekker and M. Stokhof, 403–423. Amsterdam. Institute for Logic, Language and Computation.

Morrill, Glyn V. 1993. *Type Logical Grammar: Categorial Logic of Signs.* Studies in Linguistics and Philosophy. Dordrecht, Holland: Kluwer Academic Publishers. To appear.

Pereira, Fernando C. N. 1990. Categorial Semantics and Scoping. *Computational Linguistics* 16(1):1–10.

Pereira, Fernando C. N. 1991. Semantic Interpretation as Higher-Order Deduction. In *Logics in AI: European Workshop JELIA '90*, ed. Jan van Eijck, 78–96. Amsterdam, Holland. Springer-Verlag.

Saraswat, Vijay A. 1989. *Concurrent Constraint Programming Languages.* Doctoral dissertation, Carnegie-Mellon University. Reprinted by MIT Press, Doctoral Dissertation Award and Logic Programming Series, 1993.

Scedrov, Andre. 1993. A Brief Guide to Linear Logic. In *Current Trends in Theoretical Computer Science*, ed. G. Rozenberg and A. Salomaa. World Scientific Publishing Co. Revised and expanded version of the article originally appearing in *Bulletin of the European Assoc. for Theoretical Computer Science* 41, 1990.

van Benthem, Johan. 1991. *Language in Action: Categories, Lambdas and Dynamic Logic.* Amsterdam: North-Holland.

13

Representation and Information in Dynamic Semantics

PAUL DEKKER

Among the variety of phenomena that have been the subject of study within dynamic semantics, the phenomenon of inter-sentential anaphora probably has received most attention. Traditionally, pronouns have been associated with (syntactically free, but semantically somehow bound) variables. Put in a nutshell, the semantic relationships between pronouns and their antecedents are established in a compositional way by associating both with variables, and defining the interpretation algorithm as a function updating information about the possible values of these variables. Thus, information about the values of antecedent terms is available when co-indexed pronouns are encountered.

In Dekker 1994 I have shown that the same empirical results can be obtained without labeling the subjects introduced by candidate antecedents with variables, that is, by *not* conflating natural language pronouns with a logic's variables. In the system of predicate logic with anaphora which is presented in that paper, anaphoric relationships are accounted for by keeping track of the possible values of potential antecedent terms, not of the variables they have been associated with.

In this paper I want to discuss the impact of this distinction between pronouns and variables upon the notions of representation and information involved in a dynamic semantics dealing with the interpretation of anaphoric relationships.

The research for this paper was supported by the ESPRIT Basic Research Project 6852, DYANA, Dynamic Interpretation of Natural Language. The paper is part of a more substantial contribution to DYANA deliverable R2.1.B, *Integrating Semantic theories II*, edited by Robin Cooper and Jeroen Groenendijk.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

13.1 Dynamic Semantics

In addition to the generally acknowledged context dependence of expressions, systems of dynamic semantics set out to account also for the context change potential of expressions. The leading idea is that the meaning of a sentence does not (solely) lies in its truth conditions, but rather "in the way it changes (the representation of) the information of the interpreter" (Groenendijk and Stokhof 1991). In the area of natural language semantics, a dynamic notion of meaning has been applied with considerable success in accounts of a whole array of phenomena, including presupposition, epistemic modality, the temporal structure of texts, discourse focus, defaults and, last but not least, inter-sentential anaphora.

The semantics of inter-sentential anaphoric relationships perhaps most clearly reveals the basic need of a dynamic notion of meaning. For instance, consider the following very simple sequence of sentences:

(1) A man is walking through the park. He is whistling.

These two sentences together are generally used to claim that a man who is walking through the park is whistling. That is, the pronoun *he* is interpreted as coreferential with its antecedent *a man*, even though there need not be one unique man who is walking through the park. Examples like this one pose a problem for classical, static theories of interpretation. As long as the two sentences are assigned independent denotations, there appears to be no non-ad hoc way of relating the interpretation of the pronoun with that of its antecedent.

A dynamic semantics appears to be well suited to deal with such anaphoric relationships. In a dynamic semantics the interpretation of a piece of discourse involves a constant update and adjustment of the information which is passed on for the processing of subsequent discourse. This dynamic perspective upon meaning gives us a handle to deal with an example like 1. If only, after processing the clause *a man is walking through the park*, we keep track of the possible men who walk through the park, then we are able to interpret subsequent pronouns as referring back to them.

Discourse representation theory (*DRT*, Kamp 1981, Kamp and Reyle 1993) and dynamic predicate logic (*DPL*, Groenendijk and Stokhof 1991) are two well-known examples of a dynamic semantics which deal, among others, with anaphoric relationships. In *DRT*, interpretation is in the first place defined in terms of update of (discourse) representations. These representations consist of a set of discourse referents, a kind of pegs which stand in for candidate antecedents of anaphoric pronouns, and a set of conditions on discourse referents. Anaphoric relationships are established at this level of representation by the association of pronouns with discourse referents which have been introduced by antecedent noun phrases in the current discourse representation. Discourse representations, in their turn, are interpreted in terms of the possible assignments of objects to discourse markers under which the conditions on discourse markers are satisfied. Thus, via the mediating discourse representations, pronouns and their antecedents are co-valuated.

With their system of *DPL*, Groenendijk and Stokhof show that an intermediary level of representation is not essential for an account of anaphoric relationships. It appears that the kind of information underlying the *DRT* notion of a discourse representation is in fact that of information about the values of discourse referents, or, more simply, of variables. In terms of this notion of information Groenendijk and Stokhof give a dynamic, compositional reformulation of (basic aspects of) the *DRT* account of anaphoric relationships. The *DPL* system models the interpretation of example 1 by defining interpretation as a function updating information about the values of variables. The result of interpreting the first sentence *a man is walking through the park* is an information state which contains the information that the value of a variable, say x, is a man who is walking through the park. By matching the pronoun *he* in the subsequent sentence *he whistles* with this variable x, the two terms are co-valuated.

It may be noticed that the kind of information employed in *DPL* (and implicit in *DRT*) is of a mixed syntactic/semantic nature. Information states are sets of assignments, which are functions from syntactic entities like variables to the individuals of a (predicate logic) model. (And something essentially similar goes for related approaches like those presented in Heim 1982, Barwise 1987, Rooth 1987, Zeevat 1989, Chierchia 1992, Dekker 1993, Pagin and Westerståhl 1993.) In the next two sections I will show that the *DPL* and the *DRT* interpretation procedure can be reformulated using ordinary notions of information and representation, respectively. For this, it is essential to make an explicit distinction between pronouns and variables, and to introduce anaphoric pronouns as a category of terms of its own.

13.2 Predicate Logic with Anaphora

In the *DPL* account of natural language anaphora, the information that is passed on in the process of interpretation is information about the possible values of variables that have been associated with possible antecedents of anaphoric pronouns. In Dekker 1994 I have shown that the semantics of anaphoric relationships can as well be accounted for without this intermediary use of variables. What has to be accounted for with regard to inter-sentential anaphora, is the correlation between the interpretation of pronouns and that of their antecedents. On a dynamic account of such anaphoric relationships, the possible values of (possible) antecedents must be passed on in the process of interpretation. There is no need to encode this information in terms of the values of *variables* which are associated with potential antecedents.

In Dekker 1994's Predicate Logic with Anaphora (*PLA*), the formulas of a language of predicate logic are used to represent the meanings of simple sentences or sentential clauses of natural language. Like in Groenendijk and Stokhof 1991, existentially quantified formulas are used to represent the context change potential of natural language indefinites. The difference with *DPL* is that the semantic contribution of these indefinites is not hooked up to the specific variables quantified over.

The language of this system of *PLA* is constructed like that of predicate logic from sets of relation constants R^n of arity n, from a set C of individual constants, and infinite sets V and $A = \{\mathrm{p}_i \mid i \in \mathcal{N}\}$ of variables and pronouns, respectively. A *PLA* model is an ordinary *PL* model $M = \langle D, F\rangle$, where D is a non-empty domain of individuals, and F an interpretation function which assigns individuals in D to individual constants and sets of n-tuples of individuals to n-place relation expression.

In order to deal with anaphoric relationships, *PLA* interpretation is defined as an update function on information states. These information states encode the possible values of possible antecedents of anaphoric pronouns. They consist of the 'cases' (tuples of objects) 'verifying' preceding discourse (cf., Lewis 1975). Like in Dekker 1994, I use the term 'subjects' to indicate the possible denotations of such antecedents:

Definition 1 [Information states]

- $S^n = \mathcal{P}(D^n)$ is the set of information states about n subjects
- $S = \bigcup_{n\in\mathcal{N}} S^n$ is the set of information states

A state of information about n subjects can be pictured as follows:

$$\left\{\begin{array}{c} \langle d_1,\ldots,d_n\rangle \\ \\ \vdots \\ \\ \langle d'_1,\ldots,d'_n\rangle \end{array}\right\} \qquad \begin{array}{c}\nwarrow\\ \leftarrow \\ \swarrow\end{array} \quad \begin{array}{l}\text{Possible}\\ \text{cases}\end{array}$$

$$\nwarrow \uparrow \nearrow$$

Possible values
of subjects

For a state $s \in S^n$ and $0 < j \leq n$, and for any case case $e = \langle d_1, \ldots, d_n\rangle \in s$, d_j is a possible value of the j-th subject of s, and this value will also be indicated as e_j. An information state s contains the information that the

$\imath$-th subject is a farmer who owns a donkey which is the $\jmath$-th subject, iff, for every case e in s, $e_\imath$ is a farmer, $e_\jmath$ a donkey and $e_\imath$ owns $e_\jmath$.

Before we turn to the interpretation of *PLA* formulas, I first give the *PLA* definition of information growth. In what follows, I will use the following notation conventions:

- if $e \in D^n$ and $e' \in D^m$, then $e \cdot e' = \langle e_1, \ldots, e_n, e'_1, \ldots, e'_m \rangle \in D^{n+m}$
- e' is an extension of e, $e \leq e'$, iff $\exists e''$: $e' = e \cdot e''$
- for $s \in S^n$, $N_s = n$, the number of subjects of s

Information growth consists in getting more informed about more subjects. The first aspect of information growth boils down to the reduction of the number of alternatives and the second to the extension of possible cases. Putting things together, an update of a state s is a state that consists, only, of extensions of possibilities in s:

Definition 2 [Information update] State s' is an update of state s, $s \leq s'$, iff $N_s \leq N_{s'}$, and $\forall e' \in s'\ \exists e \in s$: $e \leq e'$

As may be expected, information growth is transitive, reflexive and antisymmetric:

Observation 1 *$\langle S, \leq \rangle$ is a partial order*

We can now turn to the interpretation of terms. Constants and variables are evaluated in the usual way with respect to the model and an assignment function. The interpretation of a pronoun depends both on an information state s, and a case $e \in s$. A pronoun is associated with a subject of s, say the $\imath$-th subject, and relative to some case $e \in s$ its value is the $\imath$-th individual of e:

Definition 3 [Interpretation of terms]

- $[c]_{M,s,e,g} \quad = F(c)$ if $c \in C$
- $[x]_{M,s,e,g} \quad = g(x)$ if $x \in V$
- $[\mathrm{p}_\imath]_{M,s,e,g} \quad = e_{N_s - \imath}$ if $\mathrm{p}_\imath \in A$, $e \in s$ and $N_s > \imath$

A pronoun $\mathrm{p}_\imath$ is associated with the $i+1$-th last subject of an information state (if it exists, that is; otherwise, the interpretation of the pronoun is undefined). So, the pronoun with the index 0 picks out the subject introduced last, and its value is the last individual of any case with respect to which it is evaluated; the pronoun with index 1 picks out the forelast mentioned subject, etc., etc.

The interpretation $s[\![\phi]\!]_{M,g}$ of a *PLA* formula ϕ in an information state s is defined with respect to a model M and an assignment g. The parameters M and g are the ordinary ones from predicate logic, and they behave in the same, static, way. The state parameter s is, as it were, the dynamic one. The interpretation of a formula ϕ in a state s yields an update of s. The interpretation of *PLA* formulas is defined as follows (here, if X is a set of terms, then I_X is the smallest number greater then or equal to the index of every pronoun in X):

Definition 4 [Semantics of PLA]

- $s[\![Rt_1 \ldots t_n]\!]_{M,g} = \{e \in s \mid \langle [t_1]_{M,s,e,g}, \ldots, [t_n]_{M,s,e,g}\rangle \in F(R)\}$ (if $N_s > I_{\{t_1,\ldots,t_n\}}$)
- $s[\![t_1 = t_2]\!]_{M,g} = \{e \in s \mid [t_1]_{M,s,e,g} = [t_2]_{M,s,e,g}\}$ (if $N_s > I_{\{t_1,t_2\}}$)
- $s[\![\neg\phi]\!]_{M,g} = \{e \in s \mid \neg\exists e'\text{: } e \leq e' \ \&\ e' \in s[\![\phi]\!]_{M,g}\}$
- $s[\![\exists x\phi]\!]_{M,g} = \{e' \cdot d \mid d \in D \ \&\ e' \in s[\![\phi]\!]_{M,g[x/d]}\}$
- $s[\![\phi \wedge \psi]\!]_{M,g} = s[\![\phi]\!]_{M,g}[\![\psi]\!]_{M,g}$
- $s \models_{M,g} \phi$ iff $\forall e \in s$: $\exists e'$: $e \leq e' \ \&\ e' \in s[\![\phi]\!]_{M,g}$

An atomic formula At is evaluated with respect to any case e in the state of information s in which interpretation takes place. If such a formula only contains variables and individual constants as terms, its evaluation is in fact independent from these cases (and from s). In such a case, $s[\![At]\!]$ either equals s, iff At is classically true with respect to M and g, or $\emptyset$, iff At is classically false with respect to M and g. Only when pronouns come into play the differences between the various cases in s may matter. If, for instance, the formula is $W\mathrm{p}_i$ ("the $i+1$-th last subject walks"), then its interpretation in a state s will involve the elimination of those cases e in s of which the $i+1$-th last element does not walk (in M).

Existentially quantified formulas are taken to 'set up' discourse referents: they introduce subjects to information states. The interpretation of an existentially quantified formula $\exists x\phi$ in s involves the extension of the cases in the update of s with ϕ with x's 'witnesses'. Thus, they remain accessible for future anaphoric (co-)reference. The interpretation of a formula $\neg\phi$ in s preserves the cases in s that don't survive the update of s with ϕ, i.e., the cases that don't support ϕ. In keeping with the dynamic view on interpretation, sentence sequencing, or conjunction, involves the composition of the two update functions associated with the conjuncts. A state s supports a formula ϕ, $s \models \phi$, iff all cases in s survive the update with ϕ. In other words, ϕ is supported by s if s already contains the information conveyed by ϕ about s's subjects. (As usual, $\forall x\phi$ and $\phi \rightarrow \psi$ are taken to abbreviate $\neg\exists x\neg\phi$ and $\neg(\phi \wedge \neg\psi)$, respectively.)

Since pronouns may fail a denotation in a state s, the interpretation of a *PLA* formula can be partial. However, notice that the interpretation of a formula ψ may be partial, since it presupposes the presence of a certain number of subjects, while the interpretation of a conjunction $\phi \wedge \psi$ is total, that is, when ϕ involves the introduction of at least that number of subjects. I will say that a pronoun in a formula is resolved if it refers to a subject introduced in the very same formula, and if all the pronouns in a formula are resolved then the formula itself is called resolved.

In order to illustrate the above definitions, let us take a look at a couple of examples. First, consider the interpretation of the sentence *There is a man*, translated as $\exists x Mx$, in some state s:

(2) There is a man.

$$\begin{aligned} s[\![\exists x Mx]\!]_g &= \{e \cdot d \mid e \in s[\![Mx]\!]_{g[x/d]}\} \\ &= \{e \cdot d \mid e \in s \ \&\ d \text{ is a man}\} \ (= s') \end{aligned}$$

The interpretation of this example yields a state consisting of all the cases $e \in s$ extended with an individual d which is a man. The last subject in the resulting state (which I will refer to as s' in the following examples) simply is an arbitrary man. With a pronoun we can refer back to this man. Above I have stipulated that the i-th last subject of a state is referred back to by the pronoun with index $i-1$. So, pronoun p_0 refers to the subject introduced last, as in the following continuation of example 2:

(3) (There is a man.) He walks.

$$\begin{aligned} s'[\![W\mathrm{p}_0]\!] &= \{e' \in s' \mid \text{the last element of } e' \text{ walks}\} \\ &= \{e \cdot d \mid e \in s \ \&\ d \text{ is a man } \&\ d \text{ walks}\} \end{aligned}$$

The interpretation of this example involves the elimination of all those cases $e' \in s'$ the last element of which does not walk. The last subject in the resulting state is an arbitrary walking man.

As an illustration of the clause dealing with negation, consider a continuation of example 2 with the sentence *Nobody knows him* with translation $\neg\exists x Kx\mathrm{p}_0$:

(4) (There is a man.) Nobody knows him.

$$\begin{aligned} s'[\![\neg\exists x Kx\mathrm{p}_0]\!] &= \{e' \in s' \mid \neg\exists e''\colon e' \leq e'' \ \&\ e'' \in s'[\![\exists x Kx\mathrm{p}_0]\!]_{M,g}\} \\ &= \{e' \in s' \mid \neg\exists d'\colon e' \in s'[\![Kx\mathrm{p}_0]\!]_{M,g[x/d']}\} \\ &= \{e' \in s' \mid \neg\exists d'\colon d' \text{ knows the last element of } e'\} \\ &= \{e \cdot d \mid e \in s \ \&\ d \text{ is a man } \&\ \neg\exists d'\colon d' \text{ knows } d\} \end{aligned}$$

The last subject in this state is an arbitrary, unknown man.

To conclude this section, I point at some characteristic properties of *PLA*. Quite unlike *DPL*, *PLA* obeys the following ordinary substitution law:

Observation 2 (α-conversion) $\exists x\phi \Leftrightarrow \exists y[y/x]\phi$ *if y is free for x in ϕ and y does not occur free in ϕ*

In short, in *PLA* the ordinary notions of scope and binding apply. In fact, the subsystem of *PLA* without pronouns is fully equivalent with classical predicate logic:

Observation 3 (PL and PLA) *For any PL formula ϕ: PL $\models_{M,g} \phi$ iff $s \models_{M,g} \phi$ (for any state s)*

These two observations may go to show that the *PLA*-system is a proper extension, not modification, of ordinary predicate logic.

It is also readily established that interpretation in *PLA* always produces information update:

Observation 4 (Update) $\forall s$: $s \leq s[\![\phi]\!]$ *(if defined)*

The system of *PLA* simply models the introduction of subjects (by existentially quantified formulas) and the update of information about these subjects (by means of pronouns). It is fairly obvious that this reflects the natural language practice of indefinitely setting up and anaphorically referring back to subjects. It may be noticed that most bound variable approaches to anaphora do not have this update property. In such approaches, the introduction of a subject as the value of a variable x involves the elimination of a subject introduced earlier as the value of x. Put crudely, in a bound variable subjects may get 'dumped' as a consequence of unfortunately indexing natural language terms.

The last observation can be strengthened as follows:

Observation 5 (Registration) *For all s, $e \in D^{N_s}$: $e \in s$ & $\{e\} \models \phi$ iff $\exists e'$: $e \leq e'$ & $e' \in s[\![\phi]\!]$*

The update of a state s with ϕ contains (only) cases that register, i.e., extend, the cases in s that all by themselves support ϕ. For this reason, it is appropriate to define the notion of support in the way we did, i.e., in terms of a state s and the update of s with ϕ. I just note that such a way of defining support is crucial for an extension of the system with an account of epistemic modalities along the lines of Groenendijk et al. 1994.

13.3 Representation Theory for Anaphora

The notion of information employed in *PLA* is that of information about subjects, modeled in terms of sets of sequences of individuals. *PLA* information states are ordinary model-theoretic objects (relations), and they are the denotations of ordinary relation expressions. For this reason, the *PLA* updates of information states can also be represented as updates of relation expressions. This section presents such a representational formulation of *PLA*. The basic difference between the ensueing representation theory for anaphora (*RTA*) and *DRT* is that ordinary representations of existing formalisms are employed.

For perspicuity's sake, and in order to simplify comparison, I present a representational semantics for the *PLA*-translations of natural language sentences. The *RTA* interpretation of a *PLA* formula is defined to be a function 'updating' type theoretical relation expressions. In order to keep the correspondence with the *PLA* semantics as close as possible, I will assume Orey's relational models for the type theory, together with Muskens 1989's analysis of abstraction and application. To be precise, I use a basic type e, and relational types $\langle a_1, \ldots, a_n \rangle$, where, $a_1, \ldots, a_n$ are types, for $0 \leq n$. For each type a there is a domain D_a, defined such that $D_e \neq \emptyset$, and $D_{\langle a_1, \ldots, a_n \rangle} = \mathcal{P}(D_{a_1} \times \ldots \times D_{a_n})$. The distinguished clauses of the interpretation function are the following (for β of some type $\langle a_0, a_1, \ldots, a_n \rangle$, and γ of type $\langle a_1, \ldots, a_n \rangle$, and α and x of type a_0):

- $[\beta(\alpha)]_{M,g} = \{\langle d_1, \ldots, d_n \rangle \mid \langle [\alpha]_{M,g}, d_1, \ldots, d_n \rangle \in [\beta]_{M,g}\}$
- $[\lambda x \gamma]_{M,g} = \{\langle d_0, d_1, \ldots, d_n \rangle \mid \langle d_1, \ldots, d_n \rangle \in [\gamma]_{M,g[x/d_0]} \& d_0 \in D_{a_0}\}$

(Notice that these are harmless assumptions.) Moreover, I will assume that *PLA*'s variables are variables (of type e) of the type-theoretical language $\mathcal{L}$ and that the constants of $\mathcal{L}$ are those of *PLA*. Thus, we can assume models $M = \langle D, F \rangle$ for $\mathcal{L}$ which are also *PLA* models.

I use the following abbreviations for relational types:

- $\sigma^n \quad = \langle e_1, \ldots, e_n \rangle$
- $(a, b) = \langle a, a_1, \ldots a_n \rangle$ if $b = \langle a_1, \ldots a_n \rangle$
- $\tau^{n,m} = (\sigma^n, \sigma^m)$

The type σ^n is that of n-place relations between individuals, i.e., of states of information about n subjects. Types (a, b) mimic functional types in the relational system, and in terms of that, the type $\tau^{n,m}$ is defined to be that of functions from states of information about n subjects to states of information about m subjects. Finally, the following notation conventions will be employed. If $\vec{x}^n$ is a sequence of variables $x_1, \ldots, x_n$ (all of type e), then:

- $s(\vec{x}^n) = s(x_1) \ldots (x_n)$ (of type σ^m, for s of type σ^{n+m})
- $\lambda \vec{x}^n \gamma = \lambda x_1 \ldots \lambda x_n \gamma$ (of type σ^{n+m}, for γ of type σ^m)
- $\exists \vec{x}^n \phi = \exists x_1 \ldots \exists x_n \phi$ (of type σ^0, for ϕ of type σ^0)

For s of type σ^n, I write $\downarrow s$ for the closure $\exists \vec{x}^n s(\vec{x}^n)$ of s of type σ^0.

We may now turn to the definition of $s^n((\phi))$, the representational update of a relation term s^n of type σ^n by ϕ. The result of this, if defined, is always a relation term of some type σ^{n+m}:

Definition 5 [RTA]

- $[c]_{\vec{x}^n} = c \qquad [y]_{\vec{x}^n} = x \qquad [\mathrm{p}_i]_{\vec{x}^n} = x_{n-i}$ (if it exists)
- $s^n((Rt_1 \ldots t_m)) = \lambda \vec{x}^n\ s(\vec{x}^n) \wedge R([t_1]_{\vec{x}^n}, \ldots, [t_m]_{\vec{x}^n})$
  $s^n((t_1 = t_2)) \quad = \lambda \vec{x}^n\ s(\vec{x}^n) \wedge [t_1]_{\vec{x}^n} = [t_2]_{\vec{x}^n}$
  $s^n((\neg\phi)) \quad = \lambda \vec{x}^n\ s(\vec{x}^n) \wedge \neg\downarrow s((\phi))(\vec{x}^n)$
  $s^n((\exists y \phi)) \quad = \lambda \vec{x}^{n+m} \lambda y\ s((\phi))(\vec{x}^{n+m})$ (for $((\phi))$ of type $\tau^{n,n+m}$)

$$s^n((\phi \wedge \psi)) \quad = s((\phi))((\psi))$$

observing appropriate variable conventions (viz., that the variables in x^n and x^{n+m} are different and not free in the updated representation or in the formula inducing the update).

Let us quickly go through one example to illustrate these definitions. Consider the update of a representation s^n induced by the sequence *There is a man. He walks*, $\exists y My \wedge W\mathrm{p}_0$. First we determine the update with the first conjunct:

$$\begin{aligned} s^n((\exists y My)) &= \lambda\vec{x}^n \lambda y\ s((My))(\vec{x}^n) \\ &\Leftrightarrow \lambda\vec{x}^n \lambda y\ s(\vec{x}^n) \wedge My \end{aligned}$$

This representation, abbreviated as t^{n+1}, is next updated with the second conjunct $W\mathrm{p}_0$:

$$\begin{aligned} t^{n+1}((W\mathrm{p}_0)) &= \lambda\vec{x}^{n+1}\ t(\vec{x}^{n+1}) \wedge Wx_{n+1} \\ &\Leftrightarrow \lambda\vec{x}^n \lambda x_{n+1}\ s(\vec{x}^n) \wedge Mx_{n+1} \wedge Wx_{n+1} \end{aligned}$$

Here we see that the predicates M and W are co-instantiated. The resulting relation term denotes an information state the last subject of which is an arbitrary man who walks.

It is relatively easily established that the *RTA* update of a representation s denotes the *PLA* update of the denotation of s:

Observation 6 (RTA and PLA) $[s((\phi))]_{M,g[s/\mathbf{s}]} = \mathbf{s}[\![\phi]\!]_{M,g}$

So, as appears from this observation, *RTA* really is the representational correlate of *PLA*.

It may be clear that there is a close connection between the representations generated by *RTA*, and the *DRSs* of *DRT*. The relation terms that result from the *RTA* update of a minimal relation $\top$ (e.g., $x = x$) are easily transformed into equivalent *DRSs*. In order to be more precise, first observe that the β-normal forms of all the *RTA* representations are of the form $\lambda\vec{x}^n\ c_0 \wedge c_1 \wedge \ldots \wedge c_m$, where c_0 is $\top$, and every formula c_i, for $0 < i \leq m$, is a *condition*, viz., an atomic formula $Rt_1 \ldots t_m$ or $t_1 = t_2$, or a negation $\neg\exists\vec{y}^k\phi$ (where ϕ, in its turn, is a conjunction $c_0 \wedge c'_1 \wedge \ldots \wedge c'_l$ of conditions). These representations can be turned into *DRS*s by the following translation $()^\clubsuit$ (I assume the *DRS* syntax as presented in Groenendijk and Stokhof 1991):

- $(Rt_1 \ldots t_m)^\clubsuit = Rt_1 \ldots t_m$
  $(t_1 = t_2)^\clubsuit \quad = t_1 = t_2$
- $(\lambda\vec{x}^n\ c_0 \wedge c_1 \wedge \ldots c_m)^\clubsuit = [\vec{x}^n][(c_1)^\clubsuit \ldots (c_m)^\clubsuit]$
  $(\neg\exists\vec{y}^k\ c_0 \wedge c'_1 \wedge \ldots c'_l)^\clubsuit = \neg[\vec{y}^k][(c'_1)^\clubsuit \ldots (c'_l)^\clubsuit]$

This gives us a translation of *PLA* into the *DRS* language. The composition of i) the *RTA* update of minimal representation $\top$, ii) the β-normalization of the resulting state $()^{\beta}$, and, iii) its subsequent translation $()^{\clubsuit}$ produces a truth-conditionally equivalent *DRS*:

Observation 7 (PLA and DRT) *if* ϕ *is a resolved formula without free variables, then* $\top^0 \models_M \phi$ *in* PLA *iff* $((\top((\phi)))^{\beta})^{\clubsuit}$ *is true in* M *in* DRT

The discussion sofar shows that, simply by not conflating pronouns with variables, *DRT* and *DPL* results regarding the semantics of anaphoric relationships in natural language can be obtained using orthodox notions of representation and information. Here, one might feel a temptation to object. With the development of an *RTA*- or *PLA*-style semantics, what we have been doing is pushing the theory of interpretation to one of the two extremes of representationalism and non-representationalism. The objection then might be that this is the wrong way to go about, since, in the end, a most viable theory of interpretation should have to be located somewhere in the middle of these two extremes. Now I can agree that a comprehensive theory of natural language interpretation in the end probably is in need of informational structures in which all relevant aspects of representation and information are integrated. However, as the preceding may have shown, the phenomena we have been concerned with here appear to have no immediate bearing on *that* issue. For, as we have seen just now, the syntactic aspects of the *DPL* notion of information, or the semantic idiosyncrasies of *DRT*'s *DRSs*, simply are no inalienable ingredient of an account of the semantics of anaphoric links.

13.4 Representation and Translation

The *DPL* reformulation of *DRT* was, by and large, given in by considerations concerning the principle of compositionality, a principle which probably most would agree should be secured if it is not really too expensive. With *DPL* Groenendijk and Stokhof have developed a dynamic notion of the meaning of sentences which enables an account of local semantic dependencies between independently interpreted sentences. However, a fully general dynamic formulation of a Montagovian semantics of natural language which is also compositional at the sub-sentential level, requires a generalization of such a dynamic notion of meaning which applies to all the types that sub-sentential expressions of natural language may have. It appears that this task is complicated by the analysis of pronouns as variables. For, for a fully compositional analysis of anaphoric relationships along such lines either variables or variable assignments are to be counted in some or other way among the (model-theoretic) domains of interpretation. For this reason, in compositional elaborations additional types or sorts have been used which, by means of postulates, are made to behave like variables or variable assignments (cf., among others, Groenendijk and Stokhof 1990, Muskens 1994).

In this final section I want to show that no further syntactico-semantic assumptions really are in order if pronouns and variables are properly kept distinct. A fully compositional interpretation of a fragment of natural language, which includes anaphora, can be given in terms of an unconstrained extensional type theory, with only one basic type: that of individuals.

I now turn to an example of a fully compositional treatment of a minuscule fragment of natural language covering anaphora. The sentences of this fragment will be interpreted as *PLA* update functions on information states. *RTA* 'interpretations' are used here as the sentences' translations which denote these functions. Here we must take serious one aspect of *PLA* interpretation (and that of *RTA*, for that matter) that may have remained implicit thusfar: its polymorphism. Strictly speaking, any two *PLA* information states about a different number of subjects belong to different types. When we extend a *PLA* semantics to the sub-sentential level, this kind of type-polymorphism must be explicitly dealt with. For this reason, it is convenient to adopt a notation convention to state all the (different) translations of different types in a uniform way. I will use the following format:

$$A \rightsquigarrow B \quad [C]$$

which must be read as saying that an expression A is associated with a representation B under the condition(s) C. The condition(s) C are polymorphic type declarations. For example:

$$A \rightsquigarrow \lambda s \lambda \vec{x}^n \; s(\vec{x}^n) \wedge \phi \quad [s : \sigma^n]$$

must be taken to say that A receives all of the following translations:

1. $\lambda s \; s \wedge \phi$ (for $s : \sigma^0$)
2. $\lambda s \lambda x \; s(x) \wedge \phi$ (for $s : \sigma^1$)
3. $\lambda s \lambda x \lambda y \; s(x)(y) \wedge \phi$ (for $s : \sigma^2$)

$\vdots$

It may be noticed that, although every expression of our fragment will be associated with an (infinite) set of translations, these denote singular (polymorphic) functions.

Using this notation convention we can give the following sample translations of the basic expressions and operations of a fragment of natural language (which is left fully implicit itself):

Definition 6 [Examples of basic translations]

- *man* $\rightsquigarrow \lambda x \lambda s \lambda \vec{z}^n \; s(\vec{z}^n) \wedge \texttt{man}(x) \quad [s : \sigma^n]$
- *walk* $\rightsquigarrow \lambda x \lambda s \lambda \vec{z}^n \; s(\vec{z}^n) \wedge \texttt{walk}(x) \quad [s : \sigma^n]$
- *own* $\rightsquigarrow \lambda T \lambda x \; T(\lambda y \lambda s \lambda \vec{z}^k \; s(\vec{z}^k) \wedge \texttt{own}(y)(x))$
  $[s : \sigma^k \quad T : ((e, \tau^{k,k}), \tau^{n,m})]$

- $a \rightsquigarrow \lambda P \lambda Q \lambda s \lambda \vec{z^m} \lambda y\ (Q(y)(P(y)(s)))(\vec{z^m})$
  $[s : \sigma^n \quad P : (e, \tau^{n,k}) \quad Q : (e, \tau^{k,m})]$
- $he_i \rightsquigarrow \lambda Q \lambda s \lambda \vec{z^{n+m}}\ Q(z_{n-i})(s)(\vec{z^{n+m}})$
  $[s : \sigma^n \quad Q : (e, \tau^{n,n+m})]$

Definition 7 [Construction rules]

- $(\beta_{B/A}\ \alpha_A)_B \rightsquigarrow \beta'(\alpha') \quad [\beta' : (a, b) \quad \alpha' : a]$
- $(Not\ \pi_S)_S \rightsquigarrow \lambda s \lambda \vec{z^n}\ s(\vec{z^n}) \wedge \neg{\downarrow}\pi'(s)(\vec{x^n})$
  $s : \sigma^n \quad \pi' : \tau^{n,n+m}$
- $(\pi_S\ .\ \rho_S)_S \rightsquigarrow \lambda s\ \rho'(\pi'(s))$
  $s : \sigma^n \quad \pi' : \tau^{n,k} \quad \rho' : \tau^{k,m}$
- $(\alpha_{CN}\ who\ \beta_{IV})_{CN} \rightsquigarrow \lambda x \lambda s\ \beta'(x)(\alpha'(x)(s))$
  $s : \sigma^n \quad \alpha' : (e, \tau^{n,k}) \quad \beta' : (e, \tau^{k,m})$

In order to show how things work out, let us consider the interpretation of two examples. The first example is *A man walks. He talks*. The translation of the first sentence of this example is obtained by applying translations of *a* to fitting translations of *man* and *walk* respectively. The results of the first application can be specified as follows (all reductions here and in what follows are obtained by λ-conversion):

$$A\ man \rightsquigarrow \qquad [s : \sigma^n \quad P : (e, \tau^{n,n}) \quad Q : (e, \tau^{n,m})]$$
$$(\lambda P \lambda Q \lambda s \lambda \vec{z^m} \lambda y\ Q(y)(P(y)(s))(\vec{z^m}))(\lambda x \lambda s \lambda \vec{z^n}\ s(\vec{z^n}) \wedge \mathtt{man}(x))$$
$$\Leftrightarrow \quad \lambda Q \lambda s \lambda \vec{z^m} \lambda y\ Q(y)(\lambda \vec{z^n}\ s(\vec{z^n}) \wedge \mathtt{man}(y))(\vec{z^m})$$

Application to *walk* yields:

$$A\ man\ walks \rightsquigarrow \qquad [s : \sigma^n \quad Q : (e, \tau^{n,n})]$$
$$(\lambda Q \lambda s \lambda \vec{z^n} \lambda y Q(y)(\lambda \vec{z^n} s(\vec{z^n}) \wedge \mathtt{man}(yt))(\vec{z^n}))(\lambda x \lambda s \lambda \vec{z^n} s(\vec{z^n}) \wedge \mathtt{walk}(x))$$
$$\Leftrightarrow \quad \lambda s \lambda \vec{z^n} \lambda y\ s(\vec{z^n}) \wedge \mathtt{man}(y) \wedge \mathtt{walk}(y)$$

In a similar way we construct the translation(s) of the second sentence:

$$He_0\ talks \rightsquigarrow \qquad [s' : \sigma^m \quad Q : (e, \tau^{m,m})]$$
$$(\lambda Q \lambda s' \lambda \vec{z^m}\ Q(z_m)(s')(\vec{z^m}))(\lambda x \lambda s' \lambda \vec{z^m}\ s'(\vec{z^m}) \wedge \mathtt{talk}(x))$$
$$\Leftrightarrow \quad \lambda s' \lambda \vec{z^m}\ s'(\vec{z^m}) \wedge \mathtt{talk}(z_m)$$

Equating m with $n + 1$, we can combine the two (sets of) translations according to the construction rule for sentence conjunction. The following translation(s) result:

$$A\ man\ walks.\ He_0\ talks \rightsquigarrow \qquad [s : \sigma^n \quad s' : \sigma^m \quad \mathrm{m = n{+}1}]$$
$$\lambda s(\lambda s' \lambda \vec{z^m} s'(\vec{z^m}) \wedge \mathtt{talk}(z_m))((\lambda s \lambda \vec{z^n} \lambda y\ s(\vec{z^n}) \wedge \mathtt{man}(y) \wedge \mathtt{walk}(y))(s))$$
$$\Leftrightarrow \quad \lambda s \lambda \vec{z^n} \lambda y\ s(\vec{z^n}) \wedge \mathtt{man}(y) \wedge \mathtt{walk}(y) \wedge \mathtt{talk}(y)$$

This expression denotes the function π such that for any information state s, $\pi(s)$ is the information state that consists of all the extensions of tuples in s with a man who walks and talks.

The second example is *If a man walks, he talks.* As before, the interpretation of a conditional is stated in terms of negation and conjunction. So, *If π, then ρ* is rendered as *Not (π . not ρ).* Spelling out the construction rules, we find that:

$$(\text{If } \pi_S, \text{ then } \rho_S)_S \rightsquigarrow \qquad [s : \sigma^n \quad \pi' : \tau^{n,n+k} \quad \rho' : \tau^{n+k,n+k+m}]$$
$$\lambda s \lambda \vec{u}^n\ s(\vec{u}^n) \wedge \forall \vec{v}^k(\pi'(s)(\vec{u}^n)(\vec{v}^k) \rightarrow \exists \vec{w}^m \rho'(\pi'(s))(\vec{u}^n)(\vec{v}^k)(\vec{w}^m))$$

In this translation scheme π' can be replaced by the translation of *A man walks* given above, and ρ' by that of *He talks*. Thus, for $s : \sigma^n$, we get $\pi' : \tau^{n,n+1}$, and $\rho' : \tau^{n+1,n+1}$, and we arrive at the following specification of the translation of the conditional sentence:

$$\lambda s \lambda \vec{z}^n\ s(\vec{z}^n) \wedge \forall y((s(\vec{z}^n) \wedge \mathtt{man}(y) \wedge \mathtt{walk}(y)) \rightarrow$$
$$(s(\vec{z}^n) \wedge \mathtt{man}(y) \wedge \mathtt{walk}(y) \wedge \mathtt{talk}(y))) \Leftrightarrow$$
$$\lambda s \lambda \vec{z}^n\ s(\vec{z}^n) \wedge \forall y((\mathtt{man}(y) \wedge \mathtt{walk}(y)) \rightarrow \mathtt{talk}(y))$$

With the preceding exposition I hope to have shown that a dynamic compositional interpretation of an extensional fragment of natural language, can be appropriately stated in terms of an ordinary extensional type theory. In fact, this is something one might have expected in the first place.

13.5 Concluding Remarks

In *PLA*, the semantic connections between pronouns and their antecedents are accounted for in terms of update of information, not about the possible values of the variables associated with these antecedent terms, but about the possible values of these antecedent terms themselves. This simplified way of doing things enables us to account for the dynamics of establishing anaphoric relationships by means of a proper extension, not modification, of ordinary logical systems. Moreover, it avoids certain complications which pertain to a *DPL*-style approach to natural language anaphora. In particular, it does not automatically lead to arbitrary downdate of information.

As we have seen in this paper, the system of *PLA* also shows that a compositional treatment of the semantics of anaphoric relationships does not enforce upon us an idiosyncratic representation language, or some specialized intensional or many-sorted logic. The system of *PLA* and its representational correlate *RTA* remain well within the bounds of ordinary extensional type theory.

References

Barwise, Jon. 1987. Noun Phrases, Generalized Quantifiers and Anaphora. In *Generalized Quantifiers: Linguistic and Logical Approaches*, ed. Peter Gärdenfors. 1–29. Dordrecht: Reidel.

Chierchia, Gennaro. 1992. Anaphora and Dynamic Binding. *Linguistics and Philosophy* 15:111–83.

Dekker, Paul. 1993. *Transsentential Meditations. Ups and downs in dynamic semantics.* Doctoral dissertation, ILLC, Amsterdam.

Dekker, Paul. 1994. Predicate Logic with Anaphora (seven inch version). In *Proceedings of SALT IV*, ed. Mandy Harvey and Lynn Santelmann. Ithaca, NY. Cornell University.

Groenendijk, Jeroen, and Martin Stokhof. 1990. Dynamic Montague Grammar. In *Papers from The Second Symposium on Logic and Language*, ed. László Kálmán and László Pólos. 3–48. Budapest: Akadémiai Kiadó.

Groenendijk, Jeroen, and Martin Stokhof. 1991. Dynamic Predicate Logic. *Linguistics and Philosophy* 14:39–100.

Groenendijk, Jeroen, Martin Stokhof, and Frank Veltman. 1994. Coreference and Modality. In *Integrating Semantic Theories II*, ed. Robin Cooper and Jeroen Groenendijk. Amsterdam: ILLC. DYANA-2 deliverable R2.1.B.

Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases.* Doctoral dissertation, University of Massachusetts, Amherst.

Kamp, Hans. 1981. A Theory of Truth and Semantic Representation. In *Formal Methods in the Study of Language*, ed. Jeroen Groenendijk, Theo Janssen, and Martin Stokhof. 277–322. Amsterdam: Mathematical Centre.

Kamp, Hans, and Uwe Reyle. 1993. *From Discourse to Logic.* Dordrecht: Kluwer.

Lewis, David. 1975. Adverbs of Quantification. In *Formal Semantics of Natural Language*, ed. Edward L. Keenan. 3–15. Cambridge, U.K.: Cambridge University Press.

Muskens, Reinhard. 1989. A Relational Formulation of the Theory of Types. *Linguistics and Philosophy* 12:325–46.

Muskens, Reinhard. 1994. Compositional Discourse Representation Theory. In *Proceedings of the ninth Amsterdam colloquium*, ed. Paul Dekker and Martin Stokhof, 467–86. Amsterdam. ILLC.

Pagin, Peter, and Dag Westerståhl. 1993. Predicate Logic with Flexibly Binding Operators and Natural Language Semantics. *Journal of Logic, Language and Information* 2:89–128.

Rooth, Mats. 1987. NP Interpretation in Montague Grammar, File Change Semantics and Situation Semantics. In *Generalized Quantifiers*, ed. Peter Gärdenfors. 237–69. Dordrecht: Kluwer.

Zeevat, Henk. 1989. A Compositional Approach to Discourse Representation Theory. *Linguistics and Philosophy* 12:95–131.

14

A Persistent Notion of Truth in Dynamic Semantics

TIM FERNANDO

Abstract

For a certain interpretation of first-order formulas A as binary (input/output) relations $[\![A]\!]$ on a set S of states (specifying state transitions induced by A), a notion $|A| \subseteq S$ of truth for A is investigated, arising from a so-called double (intuitionistic) negation translation of the domain of $[\![A]\!]$ (or of, equivalently, as it turns out, the fixed points of $[\![A]\!]$). A global, Boolean-valued analysis is presented alongside a local, three-valued non-compositional approximation of it. Complications with existence and identity are exposed, and suitable generic models constructed. The analysis is carried over to a determinization of transitions between "disjunctive" sets of output states.

Introduction

The essential idea behind "dynamic semantics" (e.g., Kamp 1981) is that the semantic content of a formula (or statement) lies not so much in its truth relative to a state (or context) but in the change in state that it induces (that change being the reason the statement is *used*) That is, over a given set S of states, a formula A is interpreted as a binary relation $[\![A]\!] \subseteq S \times S$ describing transitions between states s and s'

$$s[\![A]\!]s' \quad \text{iff} \quad \text{on input } s, A \text{ can output } s' .$$

Although the present paper was completed in Stuttgart, much of the work was carried out while I was a visiting researcher at ITK, Tilburg, where I profited from discussions with Emiel Krahmer I also thank David Israel and Fernando Pereira for comments following my talk on this subject at the fourth conference on situation theory and its applications

Logic, Language and Computation
Jerry Seligmann and Dag Westerståhl, eds

By contrast, let us understand the truth value $|A|$ of a formula A to be the subset $\{s \in S : A \text{ is true at } s\}$ of S, given some definition of

$$(*) \qquad A \text{ is true at } s \ .$$

The problem addressed by the present work is to define a notion $(*)$ of truth, given a relational interpretation $[\![A]\!]$ of formulas A.

Related Work

Two obvious candidates for $(*)$ mentioned in van Benthem 1991 amount essentially to what are called acceptance and acceptability in Veltman 1990

(i) A is *accepted at* s if $s[\![A]\!]s$ (i.e., $s \in fix([\![A]\!])$)

(ii) A is *acceptable at* s if there is some state s' for which $s[\![A]\!]s'$ (i.e., $s \in dom([\![A]\!])$).

Neither definition mentions any notion of partiality on states, measuring say information content — a feature that is particularly prominent in Veltman 1990, where a state is formed from a set of worlds. If the partiality of a state s is understood to refer to the possibility of a formula A as well as its negation being (separately) acceptable in s, then adopting (ii) for $(*)$ yields an incoherent notion of truth. Option (i) would, on the other hand, mean that asserting a truth can have *no* effect on the initial state, thereby building a certain logical omniscience into states, as well as trivializing the informational significance of truths.

Proceeding more concretely, consider, for example, *Dynamic Predicate Logic* (DPL, Groenendijk and Stokhof 1991), where a state is a function from a set of variables to the universe of some first-order model. It is easy to see (under definitions reviewed in §1.1 below) that over a first-order model M where the unary relation symbol R has a non-empty denotation R^M, and a function f that *fails* to map the variable x to an element of R^M,

$$\begin{array}{ll} R(x) & \text{is neither accepted nor acceptable at } f \ , \\ \exists x\ R(x) & \text{is not accepted but is acceptable at } f \ , \end{array}$$

and

$$(\exists x\ R(x))\ \&\ R(x) \quad \text{is not accepted but is acceptable at } f \ .$$

Hence, if the formula $\exists x R(x)$ ought to be true at f, then stipulating that A is true at s precisely when A is accepted at s is too strong to account for the dynamic effects of quantification. As for identifying truth with acceptability, it is slightly embarrassing to assert that the conjunction of a true formula with an untrue formula is true. It is only slightly embarrassing in view of the somewhat novel "dynamic" interpretation of conjunction as relational composition. Even so, however, when breaking new ground, there is every reason to pause before throwing out hard-won fruits of past labor, among which one might count a Boolean-valued notion of truth.

Bringing into the picture a pre-order $\sqsubseteq$ on the set S of states (comparing information content), a natural intuition is that if $s \sqsubseteq s'$ then every formula true at s is also true at s'. That is, the truth set $|A|$ of a formula A ought to be persistent, where a set $U \subseteq S$ of states is understood to be *persistent* if for all $s \in U$, $s' \sqsupseteq s$ implies $s' \in U$. As there is no reason to expect that $fix([\![A]\!])$ or $dom([\![A]\!])$ should be persistent relative to natural measures $\sqsubseteq$ of information content (a point confirmed in section 3 below), a bit more work needs to be put into the definition of truth. And also, as it turns out, into the relational interpretation $[\![A]\!]$ of formulas A, which will be defined below to be more faithful to Kamp 1981, extended with witness constructs $\epsilon x{:}A$ for explicit quantification $\exists x\ A$.

Summary of Present Work

Building on the idea that a state $s \in S$ is *encompassed by* a set $U \subseteq S$ of states if

$$(\forall s' \sqsupseteq s)\ (\exists s'' \sqsupseteq s') \qquad s'' \in U\ ,$$

a formula A is defined to be *true at* s if s is encompassed by $fix([\![A]\!])$ — i.e.,

$$(\forall s' \sqsupseteq s)\ (\exists s'' \sqsupseteq s') \qquad s''[\![A]\!]s''\ .$$

Then the truth set $|A| = \{s \in S : A \text{ is true at } s\}$ is manifestly persistent, and is, in fact, a persistent negation applied twice. That is, $|A|$ is the image $\jmath(fix([\![A]\!]))$ of $fix([\![A]\!])$ under the the so-called *double negation translation*[1] $\jmath$ that maps subsets U of S into persistent subsets of S according to

$$\jmath(U) \quad = \quad \{s \in S\ :\ s \text{ is encompassed by } U\}\ .$$

[1] "Negation" here refers to Kripke's semantic interpretation of the intuitionistic negation $\overline{A}$ of a formula A under a forcing relation $\Vdash$ between states s partially ordered by $\leq$ and formulas

$$s \Vdash \overline{A} \qquad \text{iff} \qquad \text{not } (\exists s' \geq s)\ s' \Vdash A$$

The double negation translation $\overline{\overline{A}}$ of a formula A then yields a so-called *weak forcing* relation $\Vdash^w$ (e g , Keisler 1973)

$$\begin{aligned} s \Vdash^w A \quad & \text{iff} \quad s \Vdash \overline{\overline{A}} \\ & \text{iff} \quad (\forall s' \geq s)(\exists s'' \geq s')\ s'' \Vdash A \end{aligned}$$

typically subject to Boolean logic In the present context, the terms "weak acceptance" and "weak acceptability" are somewhat inappropriate because neither acceptance nor acceptability is persistent, as pointed out in section 3, and thus neither implies truth

A certain interpretation $[\![A]\!]$ of first-order formulas A as binary relations on finite functions from a set Var of variables to the universe $|M|$ of a first-order model M is defined supporting the following facts.

(i) Although acceptance is, as it stands, an inappropriate notion of truth for the dynamic effects introduced to handle anaphora, $\jmath$ makes acceptance acceptable in the sense that

$$\jmath(fix([\![A]\!])) \quad = \quad \jmath(dom([\![A]\!])) \ .$$

(ii) $|A|$ is compositional (relative to the logical connectives), and is moreover Boolean-valued (e.g., $|A \& B| = |A| \wedge |B|$).

(iii) The truth of A at s is determined by the behavior of A at completions of s, or more precisely, the "generic models" of s

$$A \text{ is true at } s \quad \text{iff} \quad \text{every generic model of } s \text{ satisfies } A \ .$$

(iv) $|A|$ supports truth gaps implicated in presupposition failure.

(v) Under a passage to a richer notion of ("information") state (compatible with Veltman 1990), $\jmath$ is well-behaved.

The relational interpretation $[\![A]\!]$ is essentially the part of Kamp 1981 that was meant to be reformulated by DPL, plus certain constructs introduced to handle complications with existence and identity. These constructs are reminiscent of Hilbert's epsilon terms as well as restricted parameters in situation theory.

14.1 Preliminaries

The linguistic motivation for the present work can be traced back to *Discourse Representation Theory* (DRT, Kamp 1981). Although DRT covers a good deal of ground (see, for example, Kamp and Reyle 1993), only a small part of it need concern us here. That part is henceforth referred to as DRT/DPL, having been analyzed in DPL, modulo certain simplifications that unfortunately constitute significant differences from DRT. Nonetheless, it is instructive to reverse the historical order, and present DPL (in its simplicity) before DRT/DPL. Towards that end, fix a signature **L** with equality, and a countable set Var of variables.

14.1.1 Dynamic Predicate Logic

At the heart of DPL is a translation $\cdot^{DPL}$ of first-order **L**-formulas with variables from Var into programs from (quantified) dynamic logic (see, for instance, Harel 1984) according to

$$\begin{aligned} A^{DPL} &= A? && \text{for atomic formulas } A \\ (\neg A)^{DPL} &= \neg(A^{DPL}) \\ (A \& B)^{DPL} &= A^{DPL} \ ; \ B^{DPL} \end{aligned}$$

$$(\exists x\ A)^{DPL} \quad = \quad x :=? \ ; \ A^{DPL} \ ,$$

where the programs p (in the right hand side) are interpreted relative to an L-model M as binary relations $\rho(p)$ on functions $(f, g, h, \ldots)$ from Var to the universe $|M|$ of M as follows

$$\begin{array}{rcl} f\ \rho(A?)\ g & \text{iff} & f = g \text{ and } M \models A[f] \\ f\ \rho(x :=?)\ g & \text{iff} & (\forall x' \neq x)\ f(x') = g(x') \\ f\ \rho(p;q)\ g & \text{iff} & (\exists h)\ f\ \rho(p)\ h \text{ and } h\ \rho(q)\ g \\ f\ \rho(\neg p)\ g & \text{iff} & f = g \text{ and } f \notin dom(\rho(p))\ . \end{array}$$

The reader familiar with dynamic logic may code $\neg p$ up as the test $([p]\bot)?$, where $[p]$ is the universal (box) modality labelled by the program p, and $\bot$ is some contradictory formula. As in classical logic, disjunction, implication and universal quantification can be defined in DPL from negation, conjunction and existential quantification

$$\begin{array}{rcl} A \vee B & \equiv & \neg(\neg A\ \&\ \neg B) \\ A \supset B & \equiv & \neg(A\ \&\ \neg B) \\ \forall x\ A & \equiv & \neg\exists x\ \neg A\ , \end{array}$$

where $A \equiv B$ means $\rho(A^{DPL}) = \rho(B^{DPL})$.

14.1.2 DRT/DPL

A consequence of the use of total functions as inputs and outputs in dynamic logic is that the value of a variable can be destroyed by a program interpreting quantification. One can argue, however, that there is no compelling reason for such destruction of information, and that, indeed, the use of total functions in dynamic logic obscures the expansive growth of information implicit in the introduction of discourse markers in DRT. Without getting into the details of DRT, let us switch from total functions to finite functions, and call a function from a finite subset of Var to $|M|$ an (M-)*substitution*. (We will often say "substitution" without mentioning M, with the understanding that M is fixed in the background.) The interpretations $\rho(p)$ are characterized as before (i.e., $A?$ as a test for A, sequential composition $p; q$ as relational composition $\rho(p) \circ \rho(q)$, and negation $\neg p$ as a test for the complement of the domain of $\rho(p)$), except that the relation $\rho(x :=?)$ is no longer symmetric:

$$\begin{array}{rcl} f\ \rho(x :=?)\ g & \text{iff} & dom(g) = dom(f) \cup \{x\} \text{ and} \\ & & (\forall x' \in dom(f) - \{x\})\ f(x') = g(x') \end{array}$$

for all substitutions f and g. Now, to protect a pre-existent binding on a variable $x \in Var$, let us define a *guarded assignment* $x := *$ to behave exactly like $x :=?$ on input substitutions where x is uninitialized, and to do nothing elsewhere. That is, let $x := *$ abbreviate

$$x = x? \quad + \quad \neg(x = x?) \ ; \ x :=? \ ,$$

where + is non-deterministic choice, and interpreted as union:

$$f \ \rho(p+q) \ g \quad \text{iff} \quad f \ \rho(p) \ g \ \text{ or } \ f \ \rho(q) \ g$$

for all substitutions f and g. Observe that $\neg\neg$ reduces + to DPL disjunction

$$\rho(\neg\neg(p+q)) \quad = \quad \rho(\neg(\neg p; \neg q)) \ ,$$

or, put in another way, + provides a "dynamic" alternative to DPL disjunction. Note also that because a substitution need not be defined on a variable occurring in a formula A, the program $A? + \overline{A}?$ (or $\neg(\neg A?; \neg\overline{A}?)$), where $\overline{A}$ is interpreted as the complement of A, may fail to return an output. That is, bivalence may break down because of variables undefined at the input state.

The precise relationship between the modification of DPL above and DRT is described in Fernando 1994. As detailed there, guarded assignments provide an implicit treatment of existential quantification, in the sense that the fragment of DRT covered by DPL is given by the following translation $\cdot^\bullet$ of quantifier-free first-order formulas

$$\begin{array}{rcll} (R(\overline{t}))^\bullet & = & \overline{x} := * \ ; \ R(\overline{t})? & \text{for all relation symbols } R \text{ incl. } = \\ (A\&B)^\bullet & = & A^\bullet; B^\bullet & \\ (\neg A)^\bullet & = & \neg(A^\bullet) & \end{array}$$

where the list $\overline{x} = x_1, \ldots, x_n$ in the first line for $\cdot^\bullet$ above consists of all variables occurring in $\overline{t}$, and $\overline{x} := *$ is $x_1 := *; \ldots; x_n := *$. An explicit treatment of quantification will be presented in the next section that extends this translation conservatively.

14.2 An Interpretation of Formulas as Programs

To overcome complications with identity and existence, it is useful to introduce terms that witness existential formulas $\exists x \ A$. Such terms can be viewed as "labelled variables" $y_{A,x}$ or as epsilon terms $\epsilon x : A$, similar to those of Hilbert mentioned in Israel 1994 (except that here such terms will not always be defined, thus blocking the treatment of universal quantification intended by Hilbert). (These terms can also be viewed situation-theoretically as restricted parameters; see Cooper 1992.) We will use the notation $y_{A,x}$ and $\epsilon x{:}A$ interchangably.

14.2.1 A Language

To avoid pesky variable clashes, let us be careful to partition the set *Var* of variables into two disjoint sets X and Y, where X is a countable set of "unlabelled variables" given at the outset (along with a fixed signature $\mathbf{L}$)

and the set Y of fresh "labelled" variables is defined simultaneously with the set $\mathbf{L}(Var)$ of formulas as follows.

(i) If R is an n-ary relation symbol of $\mathbf{L}$ (including $=$), and $\bar{t}$ is a list of n $\mathbf{L}$-terms with variables from $X \cup Y$, then $R(\bar{t})$ belongs to $\mathbf{L}(Var)$.

(ii) If $x \in X$, and A and B belong to $\mathbf{L}(Var)$, then $A\&B$, $A \vee B$, $\neg A$, and $\exists x\ A$ belong to $\mathbf{L}(Var)$.

(iii) If $x \in X$ and $A \in \mathbf{L}(Var)$, then the fresh variable $y_{A,x}$ (= epsilon term $\epsilon x{:}A$) belongs to Y.

Observe that $\mathbf{L}(Var)$ is *not* closed under $\supset$ or $\forall$. The expressions $A \supset B$ and $\forall x A$ can, if the reader insists, be understood as abbreviations of $\neg(A\&\neg B)$ and $\neg\exists x \neg A$, respectively.

14.2.2 A Translation

Next, modifying the DRT/DPL translation slightly, define a translation $\cdot^u$ from formulas $A \in \mathbf{L}(Var)$ into programs by

$$
\begin{array}{llcl}
(1) & (R(\bar{t}))^u & = & \bar{x} := * \ ; \ R(\bar{t})? \\
 & (A\&B)^u & = & A^u ; B^u \\
 & (A \vee B)^u & = & A^u + B^u \\
 & (\neg A)^u & = & \neg(A^u) \\
(2) & (\exists x\ A)^u & = & y_{A,x} := * \ ; \ A[y_{A,x}/x]^u \ ; \ \mathit{always}(A[y_{A,x}/x]) \ ,
\end{array}
$$

where the list $\bar{x}$ in (1) consists of all variables in X occurring in $\bar{t}$, the witness variable $y_{A,x}$ in (2) belongs to Y, and the program $\mathit{always}(A[y_{A,x}/x])$ will be described shortly. Observe that variables in X occuring freely in a formula A are automatically initialized in A^u (in accordance with (1)), whereas other variables (in Y) are initialized only when they are bound (according to (2)). That is, the choice between using unlabelled and labelled variables (to represent discourse markers) in a formula depends on how much one wishes to fuss over variable initializations.[2]

---

[2]For example, the donkey sentence

`If a farmer owns a donkey then he beats it`

can be translated as

$$\texttt{farmer}(x)\ \&\ \texttt{own}(x,x')\ \&\ \texttt{donkey}(x') \quad \supset \quad \texttt{beat}(x,x')$$

(because of effects on initializing unlabelled variables reminiscent of Pagin and D.Westerståhl 1993), or (among other possibilities) as

$$(\exists x)(\exists x')(\texttt{farmer}(x)\ \&\ \texttt{own}(x,x')\ \&\ \texttt{donkey}(x')) \quad \supset \quad \texttt{beat}(\epsilon x{:}\exists x' A, \epsilon x'{:}A')\ ,$$

where A is $\texttt{farmer}(x)\ \&\ \texttt{own}(x,x')\ \&\ \texttt{donkey}(x')$, A' is $A[(\epsilon x : \exists x' A)/x]$, and $B \supset C$ is understood as an abbreviation of $\neg(B\&\neg C)$. Observe that we lose the DPL

To make sure that a labelled variable $y_{A,x}$ is initialized only if it will always witness the formula, a program $always(A[y_{A,x}/x])$ is appended[3] to the translation of $\exists xA$. (For example, if A were $v = v \supset R(x)$ — i.e., $\neg(v = v \& \neg R(x))$ —, where $v \in Y$, then any object would witness $\exists xA$ at a state f where $v \notin dom(f)$.) Note that a witness to A can be spoiled only if variables mentioned in A^u are given new values. Fortunately, there are only finitely many such variables. This allows us to construct $always(A)$ as

$$always(A) \quad = \quad \neg\ (in(A)\ ;\ \neg(A^u))\ ,$$

where the program $in(A)$ is the non-deterministic choice $\sum In(A)$ of the set $In(A)$ of all "legal" initializations that have some bearing on A^u. Just what the set $In(A)$ is, we turn to next. The reader is warned that the description of $In(A)$ is annoyingly technical, and is perhaps best skipped on a first reading. (Afterall, once all free variables in A which can be initialized are initialized, the test $always(A[y_{A,x}/x])$ added to (2) can be ignored.) With this in mind, the definition of $In(A)$ is indented below to mark it from the rest of the text.

For $v \in Y$, let us write W_v for the formula that v should witness

$$W_v \quad = \quad C[v/x] \qquad \text{where } v \text{ is } y_{C,x}\ .$$

Then, given an $A \in \mathbf{L}(Var)$, define the sets $Y_A \subset Y$ and $Fmla_A \subset \mathbf{L}(Var)$ inductively by

if $v \in Y$ occurs in A or in some $B \in Fmla_A$ then v belongs to Y_A and W_v belongs to $Fmla_A$.

Proposition *For every $A \in \mathbf{L}(Var)$, $A \notin Fmla_A$ and the sets $Fmla_A$ and Y_A are finite. Moreover, for every $B \in Fmla_A$, $Fmla_B \subset Fmla_A$ and $Y_B \subseteq Y_A$.*

Proof That $A \notin Fmla_A$ follows from the inductive construction of Var and $\mathbf{L}(Var)$. The inclusions $Fmla_B \subset Fmla_A$ and $Y_B \subseteq Y_A$ for $B \in Fmla_A$ are easy consequences of the definitions of $Fmla_{(\,)}$ and $Y_{(\cdot)}$. An application of König's Lemma to the derivation trees of $Fmla_A$ and Y_A yields the finiteness of the sets $Fmla_A$ and Y_A. $\dashv$

Next, let

$$X_A \quad = \quad \{x \in X\ :\ x \text{ occurs freely in } A \text{ or in some } B \in Fmla_A\}\ ,$$

---

equivalences

$$(\exists xA)\ \&\ B \quad \equiv \quad \exists x(A\ \&\ B)$$
$$(\exists xA) \supset B \quad \equiv \quad \forall x(A \supset B)$$

because of variable labelling, although $\exists x$ still introduces a variable available for reference outside $\exists$'s usual scope.

[3]An alternative might be to initialize all variables in A that can be initialized before choosing a witness for $\exists xA$

$$(\exists x\ A)^u \quad = \quad init(A)\ ;\ y_{A,x} := *\ ;\ A[y_{A,x}/x]^u$$

(for some suitable program $init(A)$) but this possibility will not be pursued here.

and finally let $In(A)$ be the (finite) set of programs of the form

$$x_1 := * \; ; \; x_2 := * \; ; \; \cdots \; ; \; x_m := * \; ;$$
$$v_1 := * \; ; \; v_2 := * \; ; \; \cdots \; ; \; v_n := * \; ;$$
$$(W_{v_1})^u \; ; \; (W_{v_2})^u \; ; \; \cdots \; ; \; (W_{v_n})^u$$

where $x_1, \ldots, x_m$ is some list of elements from X_A (that does not necessarily include every $x \in X_A$), and (similarly) $v_1, \ldots, v_n$ is some list of elements from Y_A. (The sequence $(W_{v_1})^u \; ; \; (W_{v_2})^u \; ; \; \cdots \; ; \; (W_{v_n})^u$ is appended to verify that the initializations of labelled variables are "legal".) The astute reader concerned about the inductive character of our definition of $\cdot^u$ may draw from the previous proposition the

Corollary *The binary relation $\succ$ on $\mathbf{L}(Var)$ given by*

$$A \succ B \quad \text{iff} \quad B \in Fmla_A$$

is well-founded (and transitive), and for every $A \in \mathbf{L}(Var)$, the set $\{B : A \succ B\}$ is finite.

Having defined the translation $\cdot^u$, let us isolate the set S of *states* that will interest us

$$S \quad = \quad \{f \; : \; \text{for some } A \in \mathbf{L}(Var), \; \emptyset \; \rho(A^u) \; f\} \; .$$

For every $A \in \mathbf{L}(Var)$, define

$$[\![A]\!] \quad = \quad \rho(A^u) \text{ restricted to } S \; .$$

Notice that for every $A \in \mathbf{L}(Var)$, the program A^u is *guarded* in that all occurrences of random assignments in it are within guarded assignments.

Proposition 1 (Monotonicity) *For every guarded program p, and for all substitutions f and g, if $f \; \rho(p) \; g$ then $f \subseteq g$. In particular, for all $A \in \mathbf{L}(Var)$, $f[\![A]\!]g$ implies $f \subseteq g$.*

By contrast, under DPL, the output state of $[\![\exists x A]\!]$ may destroy a variable binding in the input state.

14.3 Truth of Formulas *qua* Programs

Let $\sqsubseteq$ be the restriction of $\subseteq$ to S. A natural property to require of a definition of "A is true at f" is that truth persist through $\sqsubseteq$-larger states g. Neither acceptance nor acceptability is persistent; indeed

$$f[\![A]\!]f \text{ and } f \sqsubseteq g \quad \text{need not imply} \quad g \in dom([\![A]\!]) \; ,$$

since a variable may be assigned a value at g that precludes the acceptability of A. (For example, take A to be $\neg y = y$, for $y \in Y$) To secure persistence, let us recall the double negation translation $\jmath$ mentioned in the introduction.

Definition A formula A of $\mathbf{L}(Var)$ is *true at* a state $f \in S$ if $f \in \jmath(fix([\![A]\!]))$ — i.e.,

$$(\forall g \sqsupseteq f) \; (\exists h \sqsupseteq g) \qquad h[\![A]\!]h \; .$$

Proposition 2 (Persistence) *If A is true at f and $f \sqsubseteq g$ then A is true at g.*

Proposition 2 is immediate from the transitivity of $\sqsupseteq$. The next lemma may seem rather technical, but will nevertheless prove important, particularly as $fix(\llbracket A \rrbracket)$ is not necessarily persistent.

Lemma 3 (Eventual stability) *Given an $A \in \mathbf{L}(Var)$ and a $g \in S$, there is a $\hat{g} \sqsupseteq g$ such that*

$$(\exists h \sqsupseteq \hat{g})\ h \in dom(\llbracket A \rrbracket) \quad \textit{implies} \quad (\forall h \sqsupseteq \hat{g})\ h \llbracket A \rrbracket h\ .$$

Proof The idea is to choose $\hat{g}$ such that $g\llbracket A' \rrbracket \hat{g}$, where A' is some tautology that translates (under $\cdot^u$) to a program initializing all variables in A^u which can be initialized (given input g). Towards this end, recall the definitions of $Fmla_A$, Y_A and X_A from §2.2, and choose (by the inductive construction of $\mathbf{L}(Var)$ along with Var) an enumeration $B_0, \ldots, B_n$ of $Fmla_A$ such that for all i and j with $0 \leq i < j \leq n$,

$$\text{if } B_j \text{ is } W_v \text{ then } v \text{ does not occur in } B_i.$$

Take A' to be

$$\bigwedge_{0 \leq j \leq m} x_j = x_j \quad \& \quad \bigwedge_{0 \leq i \leq n} (B_i \vee \neg B_i)\ ,$$

where $x_0, \ldots, x_m$ is an (any) enumeration of the set X_A of all variables in X occurring freely in A or in some $B \in Fmla_A$.[4] Let $Var_A = Y_A \cup X_A$ and F be the set of substitutions in S with domain contained in Var_A

$$F \quad = \quad \{f \in S \ : \ dom(f) \subseteq Var_A\}\ .$$

Observe that for every $\hat{g}$ such that $g\llbracket A' \rrbracket \hat{g}$, the restriction of $\hat{g}$ to Var_A is $\sqsubseteq$-maximal among all substitutions in F, yielding the desired property. $\dashv$

We can now provide three alternatives U to $fix(\llbracket A \rrbracket)$ satisfying $j(U) = j(fix(\llbracket A \rrbracket))$.

Theorem 4 *Given an $A \in \mathbf{L}(Var)$ and an $f \in S$, the following are equivalent.*

(i) *A is true at f.*
(ii) $f \in j(dom(\llbracket A \rrbracket))$.
(iii) $f \in j(\{h \in S : (\forall f' \sqsupseteq h)\ f' \llbracket A \rrbracket f'\})$.
(iv) $f \in j(\{h \in S : (\forall f' \sqsupseteq h)\ f' \in dom(\llbracket A \rrbracket)\})$.

[4]This proof exploits the "dynamic" interpretation of disjunction as union, $\llbracket A \vee B \rrbracket = \llbracket A \rrbracket \cup \llbracket B \rrbracket$. If disjunction $A \vee B$ is instead reduced "statically" to $\neg(\neg A \& \neg B)$, then the formula A' in the proof above will have to be replaced by one of the 2^{n+1} combinations of negated or unnegated B_i's that is satisfiable (i.e., returns an output).

Proof Clearly, (i) implies (ii), and (iii) implies (i). To see that (ii) implies (iii), assume (ii) and suppose $g \sqsupseteq f$. Appealing to Lemma 3, pick a $\hat{g} \sqsupseteq g$ such that (by (ii)) $(\forall f' \sqsupseteq \hat{g})\ f'[\![A]\!]f'$, thus validating (iii). Finally, note that (iii) implies (iv), while (iv) implies (ii). $\dashv$

As $dom([\![A]\!]) = dom([\![\neg\neg A]\!])$, Theorem 4 yields

Corollary 5 (Static projection) *For every formula A and state f, A is true at f iff for some B such that $[\![B]\!] = [\![\neg\neg A]\!]$, B is true at f.*

As $\neg\neg p$ never returns an output state different from the input, Corollary 5 raises the question as to whether or not the truth of a formula at a state f can be predicted from information about the truth of its (proper) parts at f. As the double negation translation j involves the totality of all states, it is hardly surprisingly that the answer, as the next subsection shows, is negative. Nevertheless, this exercise will prove useful warm-up for the global analysis that then follows.

14.3.1 A Local Three-valued Analysis of Truth

The present subsection is concerned with the question of predicting the truth of a formula A at a state f from information about the truth at the same state f of (proper) subformulas of A. Complementing the notion of truth, a notion of falsehood will be considered that builds on certain harmless assumptions, yielding a negation map $\sim$. Let **L** be a signature such that a bijective map $\hat{\cdot}$ on the relation symbols in **L** can be defined assigning to every R its "complement" $\hat{R}$ (also in **L**) with the same arity, such that $\hat{\hat{\cdot}}$ is the identity on the relation symbols of **L**. (A simple doubling of **L** will suffice.) An **L**-model M is understood always to satisfy

$$(\dagger) \qquad R^M \subseteq |M|^n - \hat{R}^M$$

for every n-ary relation symbol R in **L**, which is easy enough to express in the first-order language of **L**. Next, let $\sim$ be the map on $\mathbf{L}(Var)$ defined recursively by

$$\begin{aligned}
\sim R(\bar{t}) &= \hat{R}(\bar{t}) \\
\sim(A\&B) &= \sim A \vee \sim B \\
\sim(A \vee B) &= \sim A \ \& \sim B \\
\sim\neg A &= A \\
\sim(\exists x\ A) &= \neg\exists x\ A
\end{aligned}$$

(e.g., Nelson 1949, except that the clause for $\exists$ is modified, based on an identification of $\forall x A$ with $\neg\exists x\neg A$, as $\mathbf{L}(Var)$ is not closed under $\forall$). Because of ($\dagger$) above,

Lemma 6 *If $f[\![A]\!]f$ then not $f[\![\sim A]\!]f$.*

We can now assign a truth value $|A|_f$ to a formula $A \in \mathbf{L}(\mathit{Var})$ at a state f from the 3-element set $\{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$. Define

$|A|_f = \mathbf{t}$ if A is true at f,
A is false at f, $|A|_f = \mathbf{f}$, if $|{\sim}A| = \mathbf{t}$,

and

A is undetermined at f, $|A|_f = \mathbf{u}$, otherwise.

That is, $|A|_f = \mathbf{u}$ iff

$$(\exists g \sqsupseteq f)(\forall h \sqsupseteq g) \text{ not } h[\![A]\!]h \quad \text{and} \quad (\exists g \sqsupseteq f)(\forall h \sqsupseteq g) \text{ not } h[\![{\sim}A]\!]h\,.$$

A formula may be neither true nor false at a state not only because a variable may be uninitialized at that state, but also because the inclusion $\subseteq$ in (†) may be proper. Moreover, from the preceding lemma,

Corollary 7 (3-valuedness) *A cannot be both true and false at the same state.*

Although the truth values $\mathbf{t}$ and $\mathbf{f}$ persist, $\mathbf{u}$ need not. Holding a state f fixed, however, the definitions above yield the following truth tables, where $A \supset B$ is taken as an abbreviation of $\neg(A\&\neg B)$.

A	B	$\&$	$\sim$	$\vee$	$\supset$
t	**t**	**t**	**f**	**t**	**t**
t	**f**	**f**		**t**	**f**
t	**u**	**u**		**t**	**u**
f	**t**	**f**	**t**	**t**	**t**
f	**f**	**f**		**f**	**t**
f	**u**	**f**		**u**	**t**
u	**t**	**u**	**u**	**t**	**t**
u	**f**	**f**		**u**	$\neq$ **f**
u	**u**	$\neq$ **t**		$\neq$ **f**	$\neq$ **f**

The truth tables are strong Kleene except on certain entries where one of the subformulas has an undetermined truth value and no truth value can be predicted for the formula as a whole. For instance, if $|A|_f = |B|_f = \mathbf{u}$, then $|A \supset B|_f$ can be $\mathbf{t}$, as with

$$|x = 1 \supset x = 1|_\emptyset\ ,$$

or $\mathbf{u}$, as with

$$|x = 1 \supset x = 0|_\emptyset\ .$$

Observe that $|A \supset B|_f$ can never be $\mathbf{f}$, or else ${\sim}(A \supset B)$ would be true at f; i.e., $|A \,\&\neg B|_f = \mathbf{t}$, which is absurd since $|A|_f = \mathbf{u}$.

14.3.2 A Global Boolean-valued Analysis

The non-compositionality in §3.1 can be repaired through an analysis of truth that is global in that the relativization of the truth value $|A|_f$ to a state f is dropped. Instead, we define

$$|A| \quad = \quad \{f \in S \ : \ A \text{ is true at } f\} \quad (= \ j(fix([\![A]\!]))) \ ,$$

with the intuition (if the reader likes) that the truth value **u** breaks up into many different values lying (properly) between $\emptyset$ and S.

Rather than proceeding directly to $|A|$, let us try more slowly to relate $j(fix([\![A]\!]))$ to standard constructions in algebraic semantics. Towards this end, define

$$\begin{aligned} \Omega \quad &= \quad \{U \subseteq S \ : \ (\forall f \in U)(\forall g \sqsupseteq f)\ g \in U\} \\ \Omega_j \quad &= \quad \{j(U) \ : \ U \subseteq S\} \ , \end{aligned}$$

the first set Ω consisting of subsets of S persistent with respect to $\sqsubseteq$, and the second set Ω_j of j-sets.

Theorem (Folklore)

(a) $\langle \Omega, \cap, \cup, \rightarrow, \emptyset \rangle$ *is a complete Heyting algebra, where*

$$U \rightarrow U' \quad = \quad \{f \in S \ : \ (\forall g \sqsupseteq f) \ \text{ if } g \in U \text{ then } g \in U'\}$$

for $U, U' \in \Omega$, and least upper bounds are given by union, and greatest lower bounds by intersection.

(b) $\langle \Omega_j, \cap, \sqcup, c, S, \emptyset \rangle$ *is a complete Boolean algebra, where*

$$\begin{aligned} U \sqcup U' \quad &= \quad j(U \cup U') \\ c(U) \quad &= \quad \{f \in S \ : \ (\forall g \sqsupseteq f)\ g \notin U\} \end{aligned}$$

for U and $U' \in \Omega_j$, and

$$\begin{aligned} \coprod_{i \in I} U_i \quad &= \quad j(\bigcup_{i \in I} U_i) \\ \prod_{i \in I} U_i \quad &= \quad j(\bigcap_{i \in I} U_i) \end{aligned}$$

for $\{U_i \ : \ i \in I\} \subseteq \Omega_j$.

Next, to provide an algebraic semantics for **L**-sentences relative to a fixed **L**-model M, it is customary to extend the signature **L** to a signature $\mathbf{L}_M$ by throwing into **L** constants naming elements in $|M|$ (identifying $m \in |M|$ with its name). Such a step will not be required here, although **L** can, if the reader wishes, be assumed to have names for all objects in $|M|$. In any case, we will interpret not only **L**-sentences (without free variables), but also arbitrary formulas from $\mathbf{L}(Var)$ (possibly with free variables from Var). Or, put another way, the variables in *Var* will be treated as constants, which, as we will see, will serve as the usual witnesses.

Another point the reader familiar with algebraic semantics might raise is that the double negation translation j is commonly applied to a Heyting algebra — for example, the Heyting algebra Ω at hand. Although neither $fix([\![A]\!])$ nor $dom([\![A]\!])$ is necessarily in Ω, Theorem 4 suggests defining a (persistent) forcing relation $\Vdash$ between $f \in S$ and $A \in \mathbf{L}(Var)$ by either

$$f \Vdash A \quad \text{iff} \quad (\forall g \sqsupseteq f)\ g \in dom([\![A]\!])\ ,$$

or (non-equivalently)

$$f \Vdash A \quad \text{iff} \quad (\forall g \sqsupseteq f)\ g[\![A]\!]g\ .$$

Unfortunately, neither alternative provides (for instance) a decomposition of $\{f : f \Vdash A \vee B\}$ as the union (the join in the Heyting algebra Ω) of $\{f : f \Vdash A\}$ and of $\{f : f \Vdash B\}$. (Consider the case where A is $(\epsilon x\!:\!C) = (\epsilon x\!:\!C)$ and B is $\neg A$.) Some Heyting algebra other than Ω is called for; but the only natural one the present author has found is the Boolean algebra Ω_j of j-sets!

A final point before presenting our Boolean-valued analysis of truth concerns our treatment of quantification. Following our coding of guarded assignments in terms of equality, the range of quantification is restricted to objects m that "exist" in the sense that $m = m$. Also, note that for every family $\mathcal{U} \subseteq \Omega_j$, if $\bigcup\mathcal{U} \in \Omega_j$ (i.e., $\bigcup\mathcal{U} = j(\bigcup\mathcal{U})$), then $\bigcup\mathcal{U} = \coprod\mathcal{U}$.

Theorem 8 (Boolean-valued semantics) *For all $A, B \in \mathbf{L}(Var)$,*

$$\begin{aligned}
|A\&B| &= |A| \cap |B| \\
|A \vee B| &= |A| \sqcup |B| \quad (= j(|A| \cup |B|)) \\
|\neg A| &= c(|A|) \\
&= j(S - |A|) \\
|\exists x\ A| &= |(\epsilon x\!:\!A) = (\epsilon x\!:\!A)| \\
&= \bigcup \{|v = v\ \&\ A[v/x]| \ :\ v \in Var\} \\
&= \bigcup \{|A[m/x]| \ :\ m \in |M|\} \qquad \textit{if } \mathbf{L} \textit{ names all objects in } M.
\end{aligned}$$

Proof Use monotonicity (Proposition 1) and the various characterizations of $|A|$ provided by Theorem 4. For example, to see that $f \in |A\&B|$ implies $f \in |B|$, fix a $g \sqsupseteq f$ and conclude (as required by Theorem 4(ii)) that there must be some $h \sqsupseteq g$ in $dom([\![B]\!])$ since $f \in |A\&B|$. Assuming that $|A\&B| = |A| \cap |B|$ has been established, then we can appeal to static projection (Corollary 5) to reduce $|A \vee B| = |A| \sqcup |B|$ to $|\neg A| = c(|A|)$, since $[\![\neg\neg(A \vee B)]\!] = [\![\neg(\neg A\&\neg B)]\!]$. By definition, $|\neg A| = j(S - dom([\![A]\!]))$, and so by eventual stability (Lemma 3), $|\neg A| = \{f \in S : (\forall g \sqsupseteq f)\ g \notin |A|\}$. Note that eventual stability also yields $j(S - dom([\![A]\!])) = j(S - |A|)$.

Turning finally to quantification, these equations are consequences of the careful bookkeeping and coding that has gone into labelling, as well as our restriction of S to the substitutions accessible from $\emptyset$. The point is that witnesses are persistent in the sense that for all $f \in S$,

$$(\epsilon x : A) \in dom(f) \quad \text{implies} \quad f \in dom([\![\exists x\ A]\!]) \ .$$

Note the restriction in quantification to $v = v$ is required since quantified formulas translate to programs that always initialize the (translations of) variables bound by quantifiers. (Consider, for example, $\exists x\ \neg x = x$.) $\dashv$

Theorem 8 provides some evidence that our notion of truth is natural. A further argument is presented next, consisting of a construction along the partial order $\sqsubseteq$ on S, that will again force us to face the complications with existence and identity above.

14.3.3 Constructing Generic Models from States

To understand just what the double negation translation j is, we might ask what constraints on information growth from f are implied by $f \in j(U)$? Call a sequence $\{f_\alpha : \alpha < \kappa\} \subseteq S$ (for some cardinal κ) a *maximal* $\sqsubseteq$*-chain from* f if

(i) $f \sqsubseteq f_0$,
(ii) for all $\alpha < \kappa$ and $\beta < \alpha$, $f_\beta \sqsubseteq f_\alpha$, and
(iii) there is *no* $g \in S$ such that for all $\alpha < \kappa$, $f_\alpha \sqsubset g$.

Now, it is true enough that if every maximal $\sqsubseteq$-chain from f intersects U, then $f \in j(U)$. On the other hand, the converse need not hold: given a maximal $\sqsubseteq$-chain $f = f_0 \sqsubset f_1 \sqsubset f_2 \sqsubset \cdots$, the maximal chain $\{f_{2i} : i \in \omega\}$ does not intersect $U = \{f_{2i+1} : i \in \omega\}$. For a more pleasing fit with information growth, we must consider not only the partial order $\sqsubseteq$ on S, but also the formulas A in $\mathbf{L}(Var)$. In particular, we will adapt the theory of generic models described in Keisler 1973 (going back to Rasiowa and Sikorski 1963).

It will be convenient to define the basic notions from Keisler 1973 in our setting as follows. A subset U of S is a *generic set* if

(i) for all $f, f' \in U$, there is a $g \in U$ such that $f \sqsubseteq g$ and $f' \sqsubseteq g$, and
(ii) for every $A \in \mathbf{L}(Var)$, there is an $f \in U$ such that $f \in |A| \cup |\neg A|$.

Defining the *theory* $Th(U)$ *of* U to be

$$Th(U) \quad = \quad \{A \in \mathbf{L}(Var) \ : \ (\exists f \in U)\ f \in |A|\} \ ,$$

observe that condition (i) (in the definition of a generic set) is a consistency criterion, while (ii) is a completeness criterion. Next, given a signature

L, define an **L**-*free model* to be an **L**-model N, except that equality is construed as a non-logical symbol $\approx$ that is interpreted by N as some restriction of true equality

$$\approx^N \quad = \quad \{(n,n) \ : \ n \in dom(\approx^N)\} \ .$$

The term "free" here should be construed not in the usual algebraic sense but rather as in "free logic" (e.g. Bencivenga 1986), the intuition being that $dom(\approx^N)$ consists of the objects that exist. The notion of an **L**-free model N *satisfying* an **L**-sentence A, $N \models A$, is defined as in the familiar Tarskian manner (with $\approx$ regarded as a non-logical symbol), except that quantification is restricted to the domain of $\approx^N$

$$\begin{array}{lll} N \models \forall x A & \text{iff} & (\forall n \in dom(\approx^N))\ N \models A[n/x] \\ N \models \exists x A & \text{iff} & (\exists n \in dom(\approx^N))\ N \models A[n/x] \ . \end{array}$$

Now, treating Var as a set of constants, a *generic model for a* generic set U is an $\mathbf{L}(Var)$-free model N such that all its objects are named in Var

$$|N| \quad = \quad \{v^N \ : \ v \in Var\} \ ,$$

and every $\mathbf{L}(Var)$-sentence in $Th(U)$ is satisfied by N. If, moreover, $f \in U$, then N is said to be a *generic model for* f.

Theorem 9 (Generic Model Theorem) *Every $f \in S$ has a generic model; i.e., for every $f \in S$, there is a generic model for f.*

Proof The proof is similar to pp. 101 and 102 of Keisler 1973 except that (i) a generic set is constructed in a different way below (to avoid requiring that $\mathbf{L}(Var)$ be countable), (ii) conjunction, rather than disjunction, is taken to be primitive (since $\vee$ cannot be interpreted as union), and (iii) quantification must be restricted to objects n such that $n \approx n$. (Note that j turns DRT/DPL negation into intuitionistic negation.) More precisely, Theorem 9 is an immediate consequence of the following two lemmas (applied in sequence).

Lemma *Every $f \in S$ belongs to a generic set.*

Proof From an enumeration $\{A_\alpha : \alpha < \kappa\}$ of $\mathbf{L}(Var)$,construct a generic set $U = \{f_\alpha : \alpha < \kappa\}$ by taking f_0 to be f, and choosing f_α as follows, for $0 < \alpha < \kappa$. Following the notation of the proof of Lemma 3, let Var_α be the (finite) set of all variables occuring in some $B \in Fmla_{A_\alpha}$. Let g be the substitution

$$\{(v,m) \in Var_\alpha \times |M| \quad : \quad (\exists \beta < \alpha)\ v \in dom(f_\beta) \text{ and } f_\beta(v) = m\}$$

and pick a $\hat{g}$ such that $g[\![A'_\alpha]\!]\hat{g}$, where A'_α is the A' in the proof of Lemma 3, with A replaced by A_α. Set f_α to $\hat{g}$. $\dashv$

Lemma *Every generic set has a generic model.*

Proof Given a generic set U, define a (partial equivalence) relation $\equiv$ on *Var* by

$$v \equiv v' \quad \text{iff} \quad v \approx v' \in Th(U) ,$$

and form an $\mathbf{L}(Var)$-free model N by taking as its universe $|N|$ the set $\{v^{\equiv} : v \in Var\}$ of (partial equivalence) classes $v^{\equiv} = \{v' \in Var : v \equiv v'\}$ for $v \in Var$. (Note that $\emptyset \in |N|$.) Interpret the k-ary relation symbols R of $\mathbf{L}$ (inlcuding $\approx$) and the l-ary function (including constant) symbols F of $\mathbf{L} \cup Var$ according to

$$\begin{array}{lcl} R^N(v_1^{\equiv}, \ldots, v_k^{\equiv}) & \text{iff} & R(v_1, \ldots, v_k) \in Th(U) \\ F^N(v_1^{\equiv}, \ldots, v_l^{\equiv}) & = & y_{x=F(v_1,\ldots,v_l),x}{}^{\equiv} \text{ for some} \\ & & \qquad x \in X - \{v_1, \ldots, v_l\} . \end{array}$$

These are legitimate definitions (i.e., independent of the choice of representatives v of $v^{\equiv}$), since $Th(U)$ respects the laws of equality restricted to $dom(\approx^N)$ and the witnessing on variables. Notice that every term τ has a denotation τ^N, and that $\tau^N = \emptyset$ precisely if $\tau \approx \tau \notin Th(U)$. Luckily, $\emptyset \not\approx^N \emptyset$, and we have no way of expressing equality in $\mathbf{L}(Var)$ except by its approximation $\approx$.) The expected equivalence

$$N \models A \quad \text{iff} \quad A \in Th(U)$$

can be established by a routine induction on A, appealing to the consistency criterion (i) of a generic set. ⊣

⊣

Corollary 10 *Let $A \in \mathbf{L}(Var)$ and $f \in S$. A is true at f (i.e., $f \in |A|$) iff A is satisfied by every generic model for f.*

Proof As in pp. 102 and 103 of Keisler 1973, the forward direction is trivial, while the converse follows from Theorem 9. ⊣

Lastly, let us remark that the underlying $\mathbf{L}$-model M (relative to which $[\![\cdot]\!]$ and S are defined) is an elementary extension of N°, for every generic model N (based on M), where $\cdot^{\circ}$ is the obvious map from $\mathbf{L}(Var)$-free models to $\mathbf{L}$-models (obtained by throwing out all "non-existent" objects). Conversely, every countable elementary substructure of M can be obtained as such.

14.4 Discussion

A basic objection that might be raised against the double negation translation j on which our notion of truth rests is that it abstracts away the temporary instability of a formula (as Lemma 3 and ultimately Theorem 8 suggest). How is the doctrine that

the meaning of a formula is the change it induces

served by trivializing that very change? But, ideologies aside, if we refrain from reducing meaning to truth, and instead agree that truth is a "static" abstraction of meaning, then the objection loses much of its force. The question becomes, quite innocently, what does our notion of truth fail to abstract away? One answer is the partiality of $\mathbf{L}(Var)$. For although a good deal of effort was put above into developing the Boolean-valued character of the notion of truth proposed, the formally bivalent (i.e. 2-valued) interpretation in §3.3 is achieved only at the cost of a non-logical treatment of equality and a somewhat non-standard interpretation of quantification. Furthermore, as demonstrated in §3.1, one can define translations that have the effect of re-interpreting the logical connectives (for instance, negation[5]) to support some partiality. Let us conclude by briefly considering these points in relation to presuppositions, and turning to a richer notion of state similar to that in Veltman 1990 (concerning which an additional sense in which the double negation translation is persistent will be established).

14.4.1 Truth Gaps

With regard to truth gaps implicated in presupposition failure, two sources of partiality then remain:

(i) the partiality of the substitutions f, particularly relative to labelled variables (or epsilon terms), some of which cannot be initialized, and

(ii) the possibility of defining negation translations based on extension/anti-extension pairs that need not be full Boolean complements.

Comparing DRT/DPL negation $\neg$ with the negation translation $\sim$ employed in §3.1 to define falsehood, note that whereas $A \vee \neg A$ is always true, $A \vee \sim A$ may fail to be true even if for every k-ary $R \in \mathbf{L}$, $R^M = |M|^k - \hat{R}^M$ (in which case, $R^N = dom(\approx^N)^k - \hat{R}^N$ holds for every generic model N based on M). There remains the possibility of an epsilon term that can never be initialized — for example, $\epsilon x : \text{wife}(x,\text{John})$, where John has no wife (in M). Such truth gaps provide natural opportunities for giving accounts of simple (existential) cases of presupposition failure. Interpreting $\hat{R}$ to be less than the full complement of the interpretation of R offers further possibilities.

[5]Observe, however, that even though a negation $\sim$ different from DRT/DPL negation $\neg$ is introduced in §3.1, DRT/DPL negation is still useful for $\forall$ and presumably implication.

14.4.2 Information States

The analysis above is predicated on a notion of state that is somewhat impoverished, next to Heim 1983's set of "sequence-world-pairs", commonly used in accounts of presuppositions. (This is a non-trivial matter inasmuch as our notion of truth is conditioned by it.) A sequence-world-pair is simply a substitution f, coupled with the first-order model M that structures the range of f. (It becomes important to keep track of M, when a state is built from different M's.) For a sufficiently large set $\mathcal{M}$ of **L**-models, let

$$\mathcal{S} \quad = \quad \{(f, M) \ : \ M \in \mathcal{M} \text{ and } f \text{ is an } M\text{-substitution}\} \ .$$

The richer notion of a state as a subset of $\mathcal{S}$ can be obtained from the interpretation $[\![\cdot]\!]$ above, as in the classic construction of deterministic finite automata from non-deterministic ones (e.g. Hopcroft and Ullman 1979). More precisely, given an $A \in \mathbf{L}(\mathit{Var})$, lift the binary relation

$$[A] \quad = \quad \{((f, M), (g, M)) \ : \ M \in \mathcal{M} \text{ and } f[\![A]\!]_M g\}$$

on $\mathcal{S}$ to the binary relation $\mathcal{D}[A]$ on $\mathrm{Power}_+(\mathcal{S})$ ($= \mathrm{Power}(\mathcal{S}) - \{\emptyset\}$), given by

$$\mathcal{D}[A] \quad = \quad \{(U, V) \in \mathrm{Power}_+(\mathcal{S}) \times \mathrm{Power}_+(\mathcal{S}) \ : \ V = \{t \ : \ (\exists s \in U) s[A]t\}\}.$$

(Let us agree to exclude the empty set from the field of $\mathcal{D}[A]$, as that arises from the impossibility of an $[A]$-transition.) This lifting is feasible because all programs involved above are so-called distributive (i.e., definable at the level of single substitutions). We will attend later to the non-distributive programs might A of Veltman 1990. As $\mathcal{D}$ internalizes the non-determinism of $[\![A]\!]$ in states (from $\mathrm{Power}_+(\mathcal{S})$), it is natural to lift a pre-order $\leq$ on $\mathcal{S}$ comparing information content to its so-called Smyth pre-order

$$\leq_{\mathcal{D}} \quad = \quad \{(U, V) \in \mathrm{Power}_+(\mathcal{S}) \times \mathrm{Power}_+(\mathcal{S}) \quad : \quad (\forall t \in V) \ (\exists s \in U) \ s \leq t\} \ .$$

Next, let us write J for the double negation translation $j^{\sqsubseteq_{\mathcal{D}}}$ defined relative to the Smyth pre-order $\sqsubseteq_{\mathcal{D}}$ of

$$\{((f, M), (g, M)) \quad : \quad M \in \mathcal{M} \text{ and } f \sqsubseteq_M g\} \ ,$$

where $\sqsubseteq_M$ is the partial order $\sqsubseteq$ used to define truth in section 3 for a fixed **L**-model M. Finally, define an $A \in \mathbf{L}(\mathit{Var})$ to be *true at* a state $U \in \mathrm{Power}_+(\mathcal{S})$ if $U \in J(\mathit{fix}(\mathcal{D}[A]))$.

Theorem 11 Given an $A \in \mathbf{L}(\mathit{Var})$ and a $U \in \mathrm{Power}_+(\mathcal{S})$, the following are equivalent.

(i) A is true at U.
(ii) $U \in J(\mathit{dom}(\mathcal{D}[A]))$.
(iii) For every $s \in U$, A is true at $\{s\}$.
(iv) For every $s \in U$, A is true at s (according to section 4).

Proof Argue that (i) implies (ii), (ii) implies (iii), (iii) implies (iv), and finally (iv) implies (i). The first implication is immediate. The others hold because every $V \in \text{Power}_+(\mathcal{S})$ is $\sqsubseteq_{\mathcal{D}}$-dominated by some singleton (even though every singleton is also dominated by a non-singleton) and by appealing to Theorem 4. The universal quantifier in J leads to the universal (rather than existential) force in (iii) and (iv). (Observe that the theorem is quite robust under revisions of the interpretation $[\![A]\!]_M$.) $\dashv$

A final remark (which the present author plans to expand upon elsewhere) concerns Veltman 1990's program construct might, two features of which would appear to be somewhat at odds with the analysis above — viz., non-persistence and non-distributivity. Just how useful our persistent notion of truth is in analyzing the non-persistence of might remains to be seen. As for the second feature, in the same way that modality can be treated syntactically (an approach advocated in Asher and Kamp 1989 for propositional attitudes), a distributive analysis of might is possible, so long as updates are added to the underlying first-order model (exactly as sets are treated in generalized Henkin models for higher-order logic). The introduction of abstract objects into first-order models is argued and investigated at length in Asher 1993. In the case of might, the alternative syntactic approach promises to be of interest, given complications with quantification defined at the richer level of $\text{Power}_+(\mathcal{S})$-states (addressed in Groenendijk et al. 1994) — complications that the present author suspects vanish in a distributive setting.

References

Asher, N., and H. Kamp. 1989. Self-reference, attitudes, and paradox. In *Properties, Types and Meaning*, ed. G. Chierchia, B. Partee, and R. Turner. Dordrecht: Kluwer Academic Publishers.

Asher, Nicholas. 1993. *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer Academic Publishers.

Bencivenga, Ermanno. 1986. Free logics. In *Handbook of Philosophical Logic*, ed. D. Gabbay and F. Guenthner. Dordrecht: Reidel.

Cooper, Robin. 1992. Situation theoretic discourse representation theory. Manuscript, HCRC, Edinburgh.

Fernando, Tim. 1994. What is a DRS? In *Integrating Semantic Theories II*, ed. R. Cooper and J. Groenendijk. Dyana deliverable R2.1.B, 1994.

Groenendijk, J., and M. Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14.

Groenendijk, J., M. Stokhof, and F. Veltman. 1994. Coreference and modality. Hand-out for the Fourth Situation Theory and Applications Conference (Moraga, California).

Harel, David. 1984. Dynamic Logic. In *Handbook of Philosophical Logic*, ed. D. Gabbay and F. Guenthner. Dordrecht: Reidel.

Heim, Irene. 1983. On the projection problem for presuppositions. In *Proc. West Coast Conference on Formal Linguistics, 2.* Stanford: SLA: CSLI Publications.

Hopcroft, J.E., and J.D. Ullman. 1979. *Introduction to Automata Theory, Language and Computation.* Addison-Wesley.

Israel, David. 1994. The very idea of dynamic semantics. In *Proc. Ninth Amsterdam Colloquium.* Amsterdam: ILLC, University of Amsterdam.

Kamp, H., and U. Reyle. 1993. *From Discourse to Logic.* Dordrecht: Kluwer Academic Publishers.

Kamp, Hans. 1981. A theory of truth and semantic representation. In *Formal Methods in the Study of Language*, ed. T. Janssen J. Groenendijk and M. Stokhof. Amsterdam: Mathematical Centre Tracts 135.

Keisler, H. Jerome. 1973. Forcing and the omitting types theorem. In *Studies in Model Theory*, ed. M. Morley. The Mathematical Association of America.

Nelson, David. 1949. Constructible falsity. *J. Symbolic Logic* 14.

Pagin, P., and D.Westerståhl. 1993. Predicate logic with flexibly binding operators and natural language semantics. *J. Logic, Language and Information* 2(2).

Rasiowa, H., and R. Sikorski. 1963. *The Mathematics of Metamathematics.* Warsaw.

van Benthem, Johan. 1991. Logic and the flow of information. In *Proc. 9th International Congress of Logic, Methodology and Philosophy of Science*, ed. D. Prawitz et al. Amsterdam: North-Holland.

Veltman, Frank. 1990. Defaults in update semantics. In *Conditionals, Defaults and Belief Revision*, ed. H. Kamp. Edinburgh: Dyana Deliverable R2.5.A.

15

Dynamics and the Semantics of Dialogue

JONATHAN GINZBURG

Introduction

The following dialogue occurs in Harold Pinter's play *Betrayal*:

(1) a. Emma: We have a flat.
b. Robert: Ah, I see. (Pause) Nice? (Pause) A flat. It's quite well established then, your ... uh ... affair?
c. Emma: Yes.
d. Robert: How long?
e. Emma: Some time.
f. Robert: But how long exactly?
g. Emma: Five years.
h. Robert: *Five years?*
[p. 85, H. Pinter *Betrayal*, Faber, London 1991.]

Considered as an exercise in communicative interaction, this dialogue, which consists entirely of phrasal utterances from (1c) onwards, is completely effective: Robert's noting that Emma and her cohort have a flat, Emma's affirmation of the well-established nature of the affair, Robert's wondering how long the affair has been going on, Emma's informing Robert that it has gone on for five years and Robert's astonishment at Emma's informing him this, all of this which takes 50 odd words of discourse to convey, takes less than 10 words of dialogue.

Isolated from their occurrence in dialogue, however, these utterances become entirely *ineffective*: for instance, there is nothing about the expres-

Thanks to Robin Cooper, David Milward and Enric Vallduví for discussion, as well as the Natural Language Dialogue and MaC groups in Edinburgh.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

sion 'five years' that in and of itself could suggest either the resolution it gets in (1g) let alone distinguish this from the type of resolution it gets in (1h). These, then, are indications that the notion of context required for explicating dialogue must be intrinsically richer than that required for written text.

There are two aspects in particular that make dialogue an efficient medium for informational exchange. These are:

- Sharply defined *Discursive Potential*: at each point there is an extremely restricted set of topics the participants can select for discussion.

- Ellipsis: the sharply defined context severely restricts the resolution possibilities for elliptical usage.

This paper sets out to describe the nature of context needed for a semantics that captures these two features.

My starting point will be the view of context change developed in *update semantics:* in order to explicate assertion and presupposition, Stalnaker 1978 and Lewis 1979 urged construing context as a resource that represents the commonly accepted information at any given point in conversation. The initial issue will be the following: in what ways does this view need to be revised in order to accommodate the fact that **conversation involves** *discussion*?

In section 1, I emphasise the need for a structured view of context: suggesting that the *latest-move* made and the set of *questions* currently under discussion are properties of the context that require separating away from other contextual information. In section 2, I sketch what semantic properties questions are required to have in order to specify the *discursive potential*.

Even if we move towards a structured view of context, we are still some distance from engaging in a characterisation of *dialogue* since the *communicative process* is still being abstracted away from. By the communicative process I mean the fact that dialogue involves two or more distinct participants so that one participant's utterances are not automatically comprehended by the other participants. In section 3, I offer some data involving ellipsis that shows the semantic effects of the communicative process and offer a number of possible strategies towards a solution. In section 4, I sketch an analysis of the dialogue in (1) within the framework for context specified in the paper.

15.1 Discursive Potential and the Need for a Structured Context

What one might call the "classical" dynamic picture of contextual change is a view one can trace most directly to Stalnaker's 1978 paper on assertion. On such a view context at time t is identified with the set of assumptions the conversational participants hold commonly at t. This, one can assume, is also the *locus* for the variety of information assumed by situation semanticists to be present in the "discourse situation", information concerning current speaker, addressee, naming etc. How does context change? The most prominent way Stalnaker discusses is for a participant to make an assertion.[1] The descriptive content of that assertion is then added to the context *assuming none of the other participants object.* Formally speaking, nothing distinguishes this view from identifying context with the set of beliefs of a single agent ("the common ground") who from time to time receives assertoric stimuli, from Nature say, which he can either *accept* if they are not inconsistent with the existing belief set, or *reject* if they are. In what follows I often identify the Stalnakerian view (SV) with this latter conception.

The question is: what do we need to add or modify to this view in order to make it suitable for modelling dialogue? In other words, if we view dialogue as a game in which participants can pose queries and make assertions on a common *gameboard*, what attributes apart from the commonly accepted FACTS should the gameboard be specified for?

The first problematic simplification in the SV is the *Accept/Reject dichotomy*: the assumption that an assertion is either accepted, in which case propositional update occurs, or rejected, in which case propositional update is blocked. The problem with this as far as dialogue goes is that it ignores the existence of a third option: *discussion*. When one agent makes an assertion *that* p which another agent finds reason not to adopt, frequently a discussion of the question *whether* p will ensue:

(2) A: Bill left. B: Are you sure?
A: I saw his car drive away. B: That's impossible: I hear his voice upstairs. A: Look his secretary just told me he's left. ...

This brings up the more general issue that the SV does not lend itself to a characterisation of the discursive potential of a context at time t. This is true both in the uninteresting sense that, as it stands, the only dialogue move-type it can accommodate is assertion. But also true in a more profound way, which is that even if we expand the repertory of moves

[1] Of course, external factors can also change the context, for instance, as Stalnaker mentions, goats wandering uninvited into the speech location.

further, then a natural extension of the Stalnakerian view will not yield an adequate characterisation of the discursive potential at t.

In order to see this, assume we decide to add querying as an option within a Stalnakerian system. The most straightforward way to do this is to update contexts with *illocutionary* information. That is, the input to any contextual increment operation will be a proposition describing the full illocutionary information associated with the latest dialogue move:

- $ASSERT(A, p, t)$, if the move at t was an assertion by A *that p*
- $QUERY(A, q, t)$, if the move at t was a query q by A

It is clear that such an update *is* required since, assuming perfect communication, it does represent information that belongs in the common ground. The problem is this: we would now like to be in the position of being able to offer a characterisation of the possible contextual operations that can follow either an assertion or a query. But if a particular item of illocutionary information has been added in the normal way to the set of common ground FACTS, it has neither more nor less influence on what comes next than any other information in the common ground. In other words, the preconditions for the next move is the totality of what has been accepted hitherto in the common ground:

(3) a. Initial common ground: *The date today is 5 January 1995, The conversation is taking place in France, the weather outside is sunny,...*
A: Bill left yesterday.
b. New common ground: *The date today is 5 January 1995, The conversation is taking place in France, the weather outside is sunny, A asserts that Bill left yesterday,...*

To put it somewhat more figuratively: the most straightforward extension of SV to cover non-assertoric moves makes the gameboard for dialogue look like gameboards of chess or SCRABBLE. Recall that once the very initial stages have passed, a chessboard or scrabbleboard does not, for the most part, reveal to the players which moves happened when. Moreover, in SCRABBLE a player can attach his contribution adjacently to any past contribution provided there is enough room for his contribution.

Dialogue is not like SCRABBLE: an important feature of dialogue is its *locality*. Much work in Conversational Analysis has established the fundamental nature of *adjacency* in dialogue: various moves in dialogue come in move/counter-move pairs:

(4) a. A:Who left; B: Bill (query/ reply)
b. A: Open the window please! B: Sure (command/acceptance)
c. A: Hi! B: Hiya! (greeting/counter-greeting)

Any reaction to an initial member of an adjacency pair that is not an appropriate counter-move is *marked* in the sense that it will be accompanied by hesitation or hedging, unless the reaction opens a *side sequence*, roughly, a sequence of moves aimed at clarifying the content of the move that triggered the sequence:

(5) a. A: Who does Jill like? B: Jill? A: Your neighbour. B: Oh, Millie.

b. A: Put the book on the chair. B: Which chair? A: the one on the right. B: OK.

Examples such as (5a) above and their importance is discussed further in section 3.

In order to provide any kind of account of such locality phenomena we need to provide a structuring of the context. I will assume, then, that as soon as illocutionary information is *accepted*, it serves as the value of a contextual attribute which I dub: *LATEST-MOVE*. Once we allow ourselves that, we can offer an initial repertory of *reactions* to queries and assertions as following:

(6) a. If the latest move was an assertion *that* p, available moves include:

- Make a move that accepts p and update the contextual repository of facts with p.
  Or,
- Raise the issue of *whether* p as the current topic for discussion by providing information specific to *whether* p.

b. If the latest move was a query q available moves include:

- Accept q as a topic for discussion and provide information specific to q.
  Or,
- Reject q as a topic for discussion.

Are we now in a position to provide a characterisation of the *discursive potential* at t? Not quite. The problem is that while *locality* is an important feature, it is not an overriding one. Not *all* moves react to immediately preceding moves. Consider the following dialogue:

(7) A1: Why did the Pacers lose? B1: Miller wasn't playing well.
A2: Well, he scored 30 points. B2: But no offensive rebounds.
A3: hmm. Go on. B3: The refs were biassed. ...

In A1 a question is posed to which B1 responds. A2 disputes the truth of the assertion in B1. B2, on the other hand, does not dispute A2, rather tries to provide further backing for B1. The initial acknowledgement 'hmm' in A3 concedes the point. B3 returns to the issue raised in A1. A similar point emerges in (8):

(8) A1: Make a guess as to who showed up to lunch. B1: I don't want to. A2: Why not? B2: Don't want to. A3: Please. B3: Oh ok. Millie?

Here B after initially refusing to engage in a discussion of the question of who showed up for lunch finally provides B3, a response to the issue raised in A1, rather than a discussion of any of the moves B1, A2, B2, A3.

(7) and (8) show that merely structuring context with LATEST- MOVE is not sufficient for characterising the discursive potential. Rather, we also need to keep track of questions that get introduced into the context, or, as I shall put it, *that arise*, as long as they remain *under discussion.* We have already seen two important examples of questions that arise in dialogue:

- *Assert p* raises the question *whether p*
- *Query q* introduces *q*

Questions can also arise inferentially. (9) illustrates the fact that any quantificational statement changes the context in such a way as to raise the corresponding 'E-type' question:

(9) A: Several people showed up today. B: Who [are the people that showed up today]?

A fourth instance of a class of questions that arise in dialogue is the class of clarification questions illustrated above in (5).

In order to keep track of the class of questions that are (potentially) under discussion at a given point, I assume that the gameboard, in addition to FACTS and LATEST-MOVE, must also provide as value for an attribute QUD (Question Under Discussion), a partially ordered set of questions. The maximal element of QUD corresponds to the current topic of discussion.

Having motivated the tripartite view of context, I turn to a more detailed examination of the two new attributes, starting with QUD.

15.2 On the Nature of QUD

QUD is a partially ordered set of questions. To see what this amounts to and how this determines what can be discussed, I need to explain what I take questions to be and the nature of the partial ordering.

15.2.1 Questions, Propositions, and Facts in Situation Theory

The semantic framework utilized here is situation theory (e.g. Barwise and Etchemendy 1990, Barwise and Cooper 1991). The view of questions utilized here is the framework described in Ginzburg 1993. I survey the notions from that paper needed here.

Within the latter framework, the 'basic' ontology consists of a set of situations, infons and n-ary abstracts, with some algebraic structure (e.g. Barwise and Etchemendy 1990 propose the requisite structure for infons is a Heyting algebra.); a proposition $(s!\sigma)$ is constructed from a pair of s a situation, σ an infon, whereas a question $(s?\mu)$ is constructed from a pair of s a situation, μ an n-ary infon abstract.

(10) $(s!\sigma)$ is TRUE iff $s \models \sigma$. In such a case σ is a *fact*.

Questions are related to infons via two principal relations: 'ABOUT' and 'DECIDES'. Both these relations are formally characterised using the notion of informational subsumption, $\rightarrow$, within an infon algebra (Barwise and Etchemendy 1990). Thus, 'ABOUT' is a relation that, intuitively, captures the range of information associated with a question independently of factuality or level of detail:

(11) a. Jill: Is Millie leaving tomorrow? Bill: Possibly/It's unlikely/Yes/No.
b. Bill provided information about whether Millie is leaving tomorrow. (We have no indication whether this information is reliable.)

(12) a. Jill: Who is coming tonight? Bill: Millie and Chuck/Several friends of mine./Few people I know.
b. Bill provided information about who was coming that night. (We have no indication whether this information is reliable.)

The relation, which makes crucial use of the non-classicality of the infon domain (i.e. $\sigma \vee \overline{\sigma}$ is not trivial information.) is defined as follows:

(13) a. An infon τ is $ABOUT$ an abstract μ iff
$\tau \rightarrow [\bigvee(APPL-INST(\mu)) \vee \overline{\bigvee(APPL-INST(\mu))}]$
b. APPL-INST is the set of application instances of μ:
APPL-INST$(\mu) =_{def} \{\sigma | \exists f[\sigma = \mu[f]]\}$

Ginzburg 1993argues in detail for the importance of the notion of *resolvedness*, a notion of exhaustiveness relativised to the participants knowledge and purposes, for explicating the semantic properties of questions. Nonetheless, for current purposes it will suffice to consider the contextually independent notion of decidedness:

(14) A question $q = (s?\mu)$ is **decided** by a SOA τ iff
a. $s \models tau$
b. $\tau \rightarrow$ Fact-$\bigwedge(\mu)$
c. Fact-$\bigwedge_{Sit_0}(\mu) =_{def}$
$\bigwedge(\{\tau \in APPL-INST(\mu) | \exists s_0(s_0 \in Sit_0 \wedge s_0 \models \tau)\})$
if this set $\neq \emptyset$

$\bigwedge(\{\tau \in SOA_0 | \exists\sigma(\sigma \in APPL - INST(\mu) \wedge \tau = \overline{\sigma}) \wedge \exists s_0(s_0 \in Sit_0 \wedge s_0 \models \tau)\})$
Otherwise

Fact-$\bigwedge_{Sit_0}$ represents the most exhaustive application instance determined by the n-ary abstract component of a question *relative to Sit_0*. In practice, this amounts to the following:

(15) a. For yes/no questions: the factual among the two polar answers (if such exists).
b. For wh-questions: the maximal factual instantiation if such exists, otherwise, it is the negative universal quantificational answer (if such exists).

Finally, we require a notion of *dependence* between questions:

(16) a. Who committed the murder depends on who was in town.
b. q_1 DEPENDS-ON q_2 iff Any SOA τ that decides q_1 also decides q_2 (cf. Karttunen 1977)

15.2.2 Updating and Downdating QUD

Recall that when I introduced the possible reactions to assertions and queries in section 1, I presupposed the existence of a notion of an *utterance specific to a question q*. I define this as follows:

(17) Given a question $q = (s?\mu)$, a q-specific utterance is one that either:
a. Conveys information ABOUT q.
Or,
b. Conveys a question q_1 such that q DEPENDS-ON q_1.

Here, then, the notion of q-specific utterance allows in either (potential) partial answers or questions the resolution of which is a necessary condition for the resolution of q (e.g. 'A: who committed the murder? B: who was in town at the time?'.)[2]

The basic principle licensing removal of a question from QUD is the following:

(18) **QUD DOWNDATING:** If q is currently maximal in QUD, accepting information ψ that either
(a) *decides* q Or,
(b) indicates that *no information about* q can be provided
removes q from QUD and licenses adding ψ to FACTS.

The consequence of this principle is that discussion of q is licensed to continue as long as the possibility that further, new information about the question can be supplied. A3 and A4 two possible continuation alternatives

[2]This is based on Carlson 1983,p. 101.

in (19) respectively illustrate the fact that *partial answers* sometimes raise new questions but on other occasions cause a question to persist into new context. Hence, no deterministic downdating of QUD is assumed for such cases.

(19) A: Who showed up today? B: Several people.
A3: Who? A4: But can't you be more specific?

15.2.3 Order in QUD

The assumption that QUD needs to be partially ordered is motivated by the fact that more than one question can be under discussion simultaneously without conversational chaos ensuing. Although in the current paper, I take the partial order as given, a few remarks on the issue are in order. The main issue to consider is this: is the ordering *conventional* (e.g. based on order of occurrence in the dialogue) or *semantic* (e.g. via the relation'depends on' introduced above. It seems clear that in many cases a question will arise at a later point since discussing and resolving it is a precondition for (useful) discussion/resolution of an antecedently introduced question:

(20) A: Who committed the crime? B: Well, who was on parole at the time? (Implicature: 'before we start discussing your question, we need to settle this question.')

In certain cases, nonetheless, the ordering on QUD can be or needs to be negotiated:

(21) a. A: Who did Bill invite? B: Which of his friends do you know? A: Before I can answer this, you really need to answer my question. B: But I cannot answer it before you answer mine.
b. A: I'd like to ask a couple of questions. B: Shoot. A: When did you leave home? B: Uh huh. A: Where did you study French first? B: Uh huh. In which order do you want me to take these?

This argues against any syntactic principles of ordering. In fact, shows that, in principle, one effect of an utterance can be simply to effect the ordering of QUD.

15.3 Imperfect Communication and the LATEST-MOVE

15.3.1 Elliptical Follow-ups

One particularly basic and well known fact about dialogue concerns the syntax of "direct" responses: these tend to be elliptical and to display case and grammatical gender agreement with the interrogative phrase from the asked interrogative sentence. This is exemplified for German in (22):

(22) A: Wem (dative) schmeichelte der Hans?
B: Keinem (dative)/#keinen (accusative) Studenten.

In an extended version of this paper, it is shown how the potential for such ellipsis can be captured by an enrichment of QUD. Here we concentrate on a related issue: data such as the following show that elliptical contributions can arise in relation to *each constituent of the previously used utterance.* Moreover, as the data in B3 indicate, there is nothing (syntactically) "echoic" about such utterances:

(23) a. A: Jill faltered in Bibliopolis yesterday.
b. B1: Jill?/ Faltered?/In Bibliopolis?/ yesterday?
c. B2: Who?/Where?/When?/She what?
d. B3: That girl who Mary was annoying?/ In her hometown, eh?

How to explain the potential for and contribution of such utterances? I start by considering a minimalist view of the semantics of such utterances, building on Clark and Schaefer's (Clark and Schaefer 1993) work on the notion of *accepting* a dialogue contribution. I then consider what further steps need to be taken to supplement the minimalist view.

15.3.2 Clark and Schaefer

Clark and Schaefer's central point is that updating conversational common grounds *does not involve merely acceptance/rejection of contents* but also the *grounding* of any particular contribution. That is, the contributor and partners attempt to satisfy

> **The grounding criterion**: contributor and partners mutually believe that partners understood what contributor meant to a criterion sufficient for current purposes. (Clark and Schaefer 1993 p. 148.)

Clark and Schaefer suggest that in order to achieve this, dialogue moves can be partitioned into two categories: presentation moves and acceptance moves (I am slightly changing their terminology here.) A *grounded* contribution will be one that consists of a presentation move m_1 and an acceptance move $m_2(m_1)$ conditioned by m_1. While all presentation moves are, of course, overt, acceptance can be signalled implicitly, most prominently by initiation of a move directly conditioned by the content of the presentation move such as a response to a query or disagreement with an assertion.[3]

[3]Clark and Schaefer, in fact, propose a hierarchy of means among acceptance moves: a. continued attention,b. initiation of counter move (e.g. response to query), c. acknowledgement (uh huh), d. demonstration, e. display (=reprise).

The motivation for this hierarchy is to explain why dialogue doesn't get bogged down in a sequence of acceptances: given that an acceptance is also a dialogue contribution it also needs to be accepted in principle and so forth. Hence, Clark and

Clark and Schaefer, and the Conversation Analysis tradition which they build on, offer important observations and analysis which require us to make some modifications to the representation offered sofar for the values of LATEST-MOVE: once the assumption of perfect communication is dropped, it does not make much sense to think of a common contextual repository. Rather, we need to consider each participant as possessing both their own gameboard, as well as an individual mental state which provides them with stimuli from which they form beliefs about utterances that have occurred. Specifically, we need to offer the option of *accepting* an utterance, as distinct from accepting the content of an assertion or depositing a question asked in QUD. The idea, based on Cooper 1993, is to reify utterances as situations which are characterised by certain structural and semantic facts. Thus, acceptance involves acquiring a belief about the utterance, the belief that the accepter possesses the facts that she believes are mutually believed to characterize the utterance. Within the current confines of space, I will not develop the idea in formal detail, though this is undertaken in Ginzburg 1994. Informally, what this amounts to is this:

> PRE-ACCEPTANCE: If participant believes p(u) is a proposition mutually believed to exhaustively characterise the latest utterance u, she should accept u by using an acceptance move and registering LATEST-MOVE: u

15.3.3 A Simple View of Phrasal Acceptance

The notion of utterance acceptance allows us to offer a minimalist account of the meaning of one class of uses of the phrasal utterances above. In certain cases it seems reasonable to identify the meaning of utterances such as (B2) in (24) with those of (B1); namely they constitute acceptance of the utterance (A1):

(24) A1: Bill left yesterday.
B1: mmh. B2: Bill.
B3: hmm. A2: yup.

The idea is that in such cases (B2) is simply a more fleshed out version of (B1), the content of both of which can be identified as Assert(b, Accept(b, A1)), where ACCEPT is an attitude predicate describing someone in the state given in PRE-ACCEPTANCE above. This minimalist view of an utterance such as (B2), then assumes that it is not *elliptical* in the sense that it requires some fancy semantic resolution technique to provide it with

---

Schaefer posit:

> Strength of evidence principle: if e1 needed for accepting u1, and e2 for accepting e1, then e2 is weaker than e1.

a predicate. It is *context dependent*, just like (B1) is, since the context needs to provide it with an utterance of which it is a constituent and which it is being used to accept. B can then follow up on his acceptance with (B3), where the 'hmm' is already an act of slight surprise at the *proposition* communicated in A1. A2, then, is a move of accepting B's acceptance and A1 can then be regarded as grounded.

Now even this minimalist view of reprise utterances, which we might christen the *parrot* view, requires some extra bolstering since repeating just any constituent will yield wrong results:

(25) a. A1: I left yesterday.
B: You/#I, hmm.
A2: yup.
b. A1: I'll put my plate here.
B: There/# here, ok.

Hence, even if we choose to keep things simple on the semantic level as far as the "ellipsis" of phrasal utterances goes, we do need to have access to the constituency and semantics of the previous utterance. A first, informal approximation to a general acceptance move would be:

> If LATEST-MOVE: u, and u1 is a constituent of u, then if e_1 is an expression such that its content in the context of an acceptance is identical to u_1's content in u, then e_1 can be used to accept u.

Note that nothing in this rule precludes the accepting expression from being the entire previous expression and this seems reasonable enough:

(26) a. A: I'm leaving tomorrow. B: You're leaving tomorrow. Fine.
b. A: Sorry. B: Sorry. I'm not sure that's good enough.

15.3.4 Where Acceptance is not Enough

Now the minimalist, parrot view of phrasal utterances might be plausible for certain uses, for instance for examples provided in the previous section. However, it does not require much ingenuity to realize that such an analysis is not sufficient in general. Quite analogously to utterance *acceptance*, we can introduce a notion of *failure to accept* or *clarification-sought*: Whenever a conversationalist fails to believe she has been provided with the full information needed to ground the utterance, she should actively signal this. Just like 'mm' works for acceptance, there exist particles such as 'eh', 'come again' and (one type of use of) 'what' that can be used to signal complete failure to accept a given contribution. Such utterances are, in some sense, relatively uninformative since they do not pinpoint what aspect of the previous contribution was problematic. Phrasal queries are, however, a more refined means of expression. Thus, a query such as B1 in (27) signals a specific problem with the constituent 'John' of A1, not merely a failure to accept A1:

(27) A1: John left yesterday. B1: John?

Rather, it sets up the context to *discuss* a certain issue, the issue of *who A claims left yesterday.* This issue can be an unresolved one for B for a variety of reasons: it can arise because of a failure to accept the constituent 'John' of A1's contribution. It can also arise because of disagreement with A's claim, and it can arise for a combination of these reasons:

(28) (Previous dialogue continued:) A2: yup. B2: But that's impossible: I saw him an hour ago. A3: You saw John Tvetch? B3: Oh, I thought you meant John Tverry. ok.

Thus, in the particular case exemplified by B1 in (27), the utterance needn't merely raise the issue of *who A claims left yesterday*, this B could accomplish by simply saying 'who?' B1 can actually be used to ascertain whether A intended to associate with the property of *having left yesterday.* either the full *incremental* content of 'John', a particular referent, or simply its *pure* content, an existential quantifier over people named 'John'.[4]

The basic idea is, then, that in (29), both B1 and B2 shift the context to one where q_2 is maximal in QUD. In the case of B1 this is all that happens, whereas in the case of B2, a more specific question is posed in addition:

(29) a. A: John left yesterday.
B1: Who? B2: John?

b. Contextual change:
QUD: q_2 becomes maximal, where q_2: Who did A assert that left yesterday
B2 (pure content use): adds q_3 as maximal in QUD, above q_2, where q_3: Is A asserting that someone named John left yesterday ?
B2 (incremental content use): adds q_4 as maximal in QUD, above q_2, where q_4:Is A asserting that John (a particular individual) left yesterday ?

Responding to B2, A must first decide which construal to attach, pure or incremental. If the latter, a dialogue such as (28) might ensue. Whereas, if the former, her main effort will be to provide an utterance that decides q_2, for instance (30):

(30) A: You know, the guy with the broken tooth.

[4]The terms *pure/incremental content* are used in the sense of Israel and Perry 1991.

15.3.5 The Set of Questions Spawned by an Utterance

It is widely acknowledged that posing a question imposes a certain focus/ground structure on the dialogue. Consequently, short answers receive a particular construal, they are syntactically constrained to agree with the wh-role, receive phonological prominence etc. The data we have seen above suggests that putting forward *any* utterance creates the *potential* for a range of such structures. That is, each constituent not only offers its "normal" semantic contribution, but must help to build up the question obtained by abstracting over the role associated with that constituent.

There are two alternative conclusions we can draw from this phenomenon: either its set of reprise questions is actually a *product* of any utterance, or else each of the members of this set are potentially creatable from the materials deposited in the context by any particular utterance.

The first view involves wholesale complication of the content associated with each utterance: compositional construction of content is conceived of as a parallel process of building up the "normal content" simultaneously with a host of questions, at least one spawned by each constituent. This is exemplified in the following:[5]

(31) Content: Assert(A, FAIL(b,a))
Potential reprise questions: $(u?\lambda x Assert(spkr, FAIL(x, b)))$, $(u?\lambda y Assert(spkr, FAIL(j, y)))$, $(u?\lambda R Assert(spkr, R(j, b)))$,
(u is the utterance situation)

```
                                S

Content: FAIL(j,b)
Reprise Contents: FAIL(x,b), FAIL(j,y), R(j,b)

NP                                          VP
Jill                                 Content:  FAIL(x,b)

Content: j                          Reprise contents: FAIL(x,y), R(x,b)
Reprise Content: x
                                            V             NP

                                          failed          Bill

                                        Content:  FAIL     b
                                Reprise Content:  R        y
```

If we adopt this line, the information associated with a particular move will need to include the following:

[5] See Rooth 1992 for a related proposal motivated by entirely different concerns relating to focus assignment.

LATEST-MOVE: < set of reprise questions generated by u; Illocutionary/descriptive content(u) >

This first view is not particularly attractive, not least because it is an instance of a worst case analysis: some and in fact all of the reprise questions might not be raised in subsequent dialogue. An alternative view of how to specify the family of reprise questions is based on the notion of *delay*: rather than incrementally building up a class of questions, simultaneously with interpretation, we do nothing. What we do do is view the *entire utterance*, both syntactic and semantic information, as a resource that gets transmitted to the subsequent context. That this is cognitively plausible is backed by our ability to reprise accurately at least the most recent utterance (and by a host of various psychological studies going back to the classic experiments of Sachs 1967.) With this resource at hand, it is then possible to construct any reprise question pertaining to that utterance, as the need arises by backtracking to the appropriate constituent and solving the appropriate higher order unification equation (Dalrymple et al. 1991):

(32) Backtracking to the constituent 'Bill' in 'I failed Bill': Reprise question raised: $(u?\lambda y Assert(spkr, FAIL(a, y)))$
Solution to: P(b) = Assert(A, FAIL(b,a))

If we adopt this line, the information associated with a particular move will need to include the following:

LATEST-MOVE: < syntax and semantics of u >

15.4 Concluding Remarks

Starting out from a Stalnakerian view of context as set of currently held assumptions, I have provided motivation for a more structured view of context in dialogue. I have urged that a *global* view of context whereby all information is of equal strength in determining the preconditions for the next move should be supplanted by a context which keeps track of the following attributes:

- FACTS: the set of currently accepted facts.
- LATEST-MOVE represents the syntax and semantics of the *latest-move* made. It is permissible to make whatever moves are available as reactions to the latest-move.
- QUD: is a partially ordered repository that specifies the currently discussable questions. If q is maximal in QUD, it is permissible to provide information *about* q or a question q_1 on which q depends.

In order to make this concrete, let us reconsider the initial dialogue from *Betrayal* and see how the beginnings of a formal analysis can be provided within the current framework.

Emma: **We have a flat.**

Robert. (1a) **Ah, I see.** *By providing the second element of an adjacency pair, an acceptance of an assertion, Robert grounds Emma's utterance.*
(Pause) Nice? (Pause) *Robert raises an issue, presumably* whether the flat is nice, *though an account of such ellipsis is not available within the current account. The following pause indicates this utterance is not grounded by Emma. Robert moves on:*
A flat. *Robert uses this phrasal utterance to indicate recontemplation of Emma's utterance.*
It's quite well established then, your uh affair? *This question is now maximal in Robert's QUD*

Emma: **Yes.** *By providing the second element of an adjacency pair, Emma grounds Robert's utterance; she offers a response that decides the maximal question in QUD.*

Robert: **How long?** *Emma's response enabled Robert to downdate the question introduced in (1a) from QUD and update FACTS with this information. This fact raises the question:* how long has the affair been well established, *which Robert deposits in his QUD*

Emma. **Some time.***Emma grounds Robert's query; her response does not* decide *the question he posed.*

Robert: **But how long exactly?** *Robert poses a more specific question, one on which the previously maximal element of QUD depends.*

Emma: **Five years.***Emma offers a response that decides the maximal question in QUD.*

Robert: (3b) **Five years?** *Robert expresses his astonishment by an unwillingness to ground Emma's utterance: he poses a clarification query paraphrasable as:* Are you asserting that the affair has been well established for five years.

References

Barwise, Jon, and Robin Cooper 1991 Simple Situation Theory and its graphical representation In *Partial and Dynamic Semantics III*, ed. Jerry Seligman DYANA Deliverable R2.1.C. University of Edinburgh: Cemtre for Cognitive Science.

Barwise, Jon, and John Etchemendy. 1990. Information, Infons, and Inference. In *Situation Theory and Its Applications, I*, ed. Robin Cooper, Kuniaki Mukai, and John Perry. p. 33–78. CSLI Lecture Notes Number 22. Stanford: CSLI Publications.

Carlson, Lauri. 1983. *Dialogue Games.* Synthese Language Library. Dordrecht: D. Reidel.

Clark, Herb, and Edward Schaefer. 1993. Contributing to Discourse. In *Arenas of Language Use.* p. 259–94. Stanford: CSLI Publications. Reprinted.

Cooper, Robin. 1993. Integrating different information in linguistic interpretation. In *First International Conference on Linguistics at Chosun University.* p. 79–109.

Dalrymple, Mary, Fernando Pereira, and Stuart Shieber. 1991. Ellipsis and Higher Order Unification. *Linguistics and Philosophy* 14:p. 399–452.

Ginzburg, Jonathan. 1993. Resolving Questions. *HCRC research paper* RP-46.

Ginzburg, Jonathan. 1994. An Update Semantics for Dialogue. *HCRC Ms* xx.

Israel, David, and John Perry. 1991. What is Information. *CSLI Report* 91-145.

Karttunen, Lauri. 1977. Syntax and Semantics of Questions. *Linguistics and Philosophy* 1:p. 3–44.

Lewis, David K. 1979. Score Keeping in a Language Game. In *Semantics from different points of view*, ed. R. Bauerle. p. 172–187. Springer.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1:p. 75–116.

Sachs, Jacqueline. 1967. Recognition memory for syntactic and semantic aspects of connected discourse. *P&P* 2:p. 437–442.

Stalnaker, Robert C. 1978. Assertion. In *Syntax and Semantics, Volume 9*, ed. P. Cole. p. 315–332. New York: AP.

16

Towards a Channel-Theoretic Account of the Progressive

SHEILA GLASBEY

Introduction

Capturing the semantics of the progressive is not a straightforward matter. Twenty years of attempts to pin down the exact conditions under which a progressive sentence is true have not yet, we will argue below, yielded an adequate account — although substantial progress has been made.

After a brief review of some of the problems, we will look at two recent accounts of the semantics of the progressive, those of Landman (1992) and Asher (1992), and argue that neither account is completely satisfactory. We will propose a new treatment of the progressive, based upon recent developments in channel theory (Barwise and Seligman 1994). The idea is not entirely new: Hinrichs (1983) gave an account of the progressive in terms of situation theoretic constraints, and Cooper (1985) made further suggestions along these lines. Our account may be regarded as an attempt to develop Hinrichs' basic idea in the light of the more sophisticated notions of information flow given by channel theory.

We will argue that the notion of **natural regularity** embodied in channel theory is exactly what is needed to give an account of the progressive which is precisely expressed and explains the data. In particular, we will

---

This work has benefited greatly from discussions with Robin Cooper and Jerry Seligman. I would also like to thank Janet Hitzeman and members of the Tense and Aspect Group and Meaning and Computation Group at the Centre for Cognitive Science, University of Edinburgh for helpful comments and discussion. I am also grateful for the comments of participants at the Conference on Information-Oriented Approaches to Language, Logic and Computation, Saint Mary's College of California, June 1994. The work is supported by an EPSRC postdoctoral fellowship.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

emphasize the fact that an account using natural regularities allows us to separate out two features — the basic semantics of the progressive and the ways in which we may reason about the information conveyed. While both are important, we will argue that it is advantageous to keep them separate and show how this allows us to overcome problems in Asher's account, which is based on defaults.

We will then apply the account to a range of examples, including some which are handled by Landman and/or Asher and some which are not.

16.1 A Brief History of the Progressive

An early truth-conditional account of the progressive was given by Bennett and Partee (1972), in terms of interval semantics. The idea was that a progressive sentence is true at an interval t if t is part of a larger, later-ending interval t' at which the corresponding non-progressive sentence is true. The problem is that this predicts that a progressive sentence is only true if the corresponding non-progressive sentence is also true. In other words, there is no allowance for interruptions. Yet it is clearly possible for the progressive of an accomplishment[1] to be true even if the complete event never occurs — we can say, for example:

(1) Mary was building a house when she ran out of money and had to stop.

In order to get over this problem (known as the **imperfective paradox**), Dowty (1979) used **inertia worlds**. The non-progressive is no longer required to be true in the actual world, but it must be true at a larger interval in all the inertia worlds. The inertia worlds for an interval i and a world w are those worlds which are identical to w up to i, but in which from then on "nothing unexpected happens" and events take their normal, natural course. However, there is a problem in saying exactly what is meant by an inertia world. For example, it has been argued that for:

(2) Mary was crossing the road when she was hit by a cyclist.

if events take their normal, natural course then Mary will still be hit by the cyclist. It has been proposed that we remove everything "external to the event" and if this allows the complete event to be realised, then the progressive is true.[2] Landman (1992) points out a problem with this, however — that of sentences like:

---

[1] We use the terminology of Vendler 1967.

[2] Sentences like

(3) Irene is making fish stew but the cat is eating the fish.

also pose problems for Dowty's account — there are no worlds where Irene makes the soup and the cat eats the fish, because the two outcomes are incompatible. This example is credited by Asher to Irene Heim. We will show how such examples can be dealt with in our channel-theoretic analysis proposed below.

(4) Mary was wiping out the Roman army.

which seems unacceptable when used to describe a scenario where Mary, a normal human being with only conventional weapons at her disposal, has (for example) killed two Roman soldiers and is about to attack a third. The reason we judge such a progressive to be unacceptable seems to have to do with the fact that the project is absurd — we know that Mary hasn't a chance in hell of achieving her objective.

Landman proposes a way to overcome this problem. The idea is that we follow the progress of an event e (corresponding to the complete event, as described by the non-progressive). If e stops in the actual world w, we go to the closest possible world where e doesn't stop, and follow its progress there. But we can only move to the closest world if that world is what Landman calls a "reasonable option" from w, on the basis of what is "internal to e". If e is interrupted in this new world, we move to the closest world to that one, provided once again that this closest world is a reasonable option from w. This process is continued until a world is found in which the e goes to completion, or until we come to a point where going to the closest world is no longer reasonable in the above sense. If the latter happens, the progressive fails to be acceptable.

Landman is thus able to explain why (4) is no good. Suppose Mary gets killed by a soldier after the first few minutes. Then we can go to a world where everything is the same except that the soldier does not kill her. Of course, she will probably get killed very shortly by another soldier. It is arguably a reasonable option to go to a world where that soldier does not kill her, either. But there is a limit to how long we can keep on doing this. After removing a few soldiers, Landman argues, it is no longer a reasonable option to remove any more, and we are forced to abandon Mary to her fate.

The notion of **reasonable option** is not made precise. Landman might argue that this is just what we want — it gives us some flexibility in which progressives are allowed, depending on what the hearer considers to be "reasonable". This flexibility seems right (and we will build it in to our account with channels), but what is less satisfactory is that there is no explanation of where the notion of what is reasonable comes from. There is simply no way to capture formally the fact that what is reasonable to one speaker may be unreasonable to another. Neither is the notion of what counts as **internal to the event** made precise. Consider:

(5) John is walking to the shops.

Now suppose John habitually takes a walk from his house to the shops. A neighbour observes him daily setting out, and returning with his shopping. Suppose that on the way to the shops, there is a turning which leads to the park, and one fine morning John decides, as he sets out, to forgo his shopping and take a stroll in the park instead. The neighbour, watching him set out, is not aware of the change of plan, and says to herself:

(6) John is walking to the shops.

She is wrong, however — the progressive is not true on this occasion. But of course it's a very reasonable option for John to walk to the shops — he does it every day. What is different today is John's intention, which is invisible to his neighbour. How can Landman's account rule out (6) in this case? Landman could point out that the notion of reasonable option depends on what is "internal to the event". If we can argue that John's intention to go to the park is internal to the event, then we can say that on the basis of this intention it is *not* a reasonable option to go to a world where he doesn't turn off to the park, but goes to the shops. This leaves, however, the problem of how to make precise the notion of **internal to the event**.

Landman admits that pairs of progressives like the one below are problematic for his account as it stands.

(7) We were flying to Manchester.

(8) We were flying to Havana.

It appears that there are single scenarios which are correctly described by both (7) and (8). Suppose that we bought tickets to Manchester and got on the plane which took off and headed for Manchester. On the way, a band of hijackers forced the pilot to take us instead to Havana. In restrospect it seems quite possible to say:

(9) We were flying to Manchester but some hijackers forced the pilot to take us to Havana.

and also to say:

(10) We were, although we didn't know it at the time, flying to Havana.

Landman proposes extending his account to deal with such examples by incorporating some notion of **perspective**. Depending on the perspective of the speaker, one or other of the progressives may be true. However, he does not attempt to formalise this idea, and it seems that he would need to introduce a considerable amount of new machinery in order to do so. We will show below how channel theory provides us with the theoretical tools to begin to make the notion of perspective precise.

Asher (1992) gives a semantics for the progressive which employs defeasible reasoning. Use of the progressive relies on the existence of a **default** to the effect that, in the absence of any information to the contrary, we may conclude that the corresponding non-progressive sentence is true. If, however, there is information to the contrary, as is the case in:

(11) Mary was crossing the road when she was hit by a cyclist.

then the default is overridden by the more specific information to the contrary, and we may not conclude that she crossed the road.

Asher uses a possible worlds semantics which makes no use of events or event types, and thus in order to capture what he calls the "relevant features" of the progressive state[3] that participate in the default, he introduces a theoretical notion of **perspective**, which excludes those aspects of the state which are not considered relevant. Thus, which characteristics we can count or discount depends on the the perspective taken. Exactly which perspectives are licensed at a particular point of the discourse is a matter that Asher does not make precise.

There are some progressives where the default notion does not seem to be correct. One example, considered by Asher himself, is:

(12) Mary was crossing the minefield.

Here, the problem is that the progressive is judged acceptable even though there appears to be no default that people crossing minefields generally arrive at the other side. A more likely outcome (or, at least, an outcome that is equally likely) is that the person will be blown up *en route*. Now a possible rejoinder is to say that special knowledge about Mary or the minefield might allow us to use the progressive. If, for example, Mary is an expert in crossing minefields and has suitable mine-detecting equipment, the default may well be that she gets safely across. Another possibility, raised by Asher, is that (12) may be used if Mary is unaware that it's a minefield she's crossing. This might allow us, says Asher, to take a perspective which excludes the information that it is a minefield.

However, it is possible to construct examples which pose problems for both these strategies. Consider the following discourse:

(13) I looked out of the observation hut window and was horrified at what I saw. Mary was crossing the minefield. I realised, remembering her state of despair when she spoke to me the day before, that she was trying to get herself blown up.

In this scenario, Mary not only knows that it's a minefield, but is exploiting the fact — she expects to be killed. It is therefore completely unjustifiable to take a perspective which excludes the minefield. The narrator, too, is fully aware that it's a minefield. It would thus be extremely difficult to justify any perspective which left out that fact. It is hard to see how Asher's notion of default can account for this example.

Another problematic example is:

[3]Asher treats progressives are statives, following Vlach (1981) and others.

(14) "I hear that John is writing a novel. He won't finish it, of course. Hardly anyone ever does."

Here, the speaker feels able to describe what John is doing as 'writing a novel', even though she clearly perceives no default to the effect that people writing novels normally finish them. Indeed, she believes that the default goes the other way and claims that people generally don't finish novels. It does not seem possible to find any other perspective here which might license the default, as this is simply a piece of directly reported speech. So we are left with the same kind of problem as in the minefield example.

The minefield example is also problematic for Landman. In the scenario described by (12), Landman would need to explain the acceptability of the progressive by being able to move to successive closest possible worlds, removing "one mine at a time", just as in (4) we removed one Roman soldier at a time. Remember, though, that we can only remove another mine, each time, if to do so is a reasonable option from the actual world w. Now we might well argue that there comes a point, long before Mary reaches the other side, at which it is no longer a reasonable option to remove any more mines. This is certainly the case if we expect Mary not to get across, as in (12). So Landman is left without a way to explain why (12) is acceptable.

16.2 Towards an Improved Account

It seems that both Asher and Landman have succeeded in capturing something important about the semantics of the progressive, yet neither account is completely acceptable as it stands. Let us try to analyse what is going wrong, and attempt to put it right.

One possibility is that by uttering (12), the speaker is not making use of a specific default about crossing minefields in particular, but is making use of a more general default — perhaps one concerned with "crossing things" in general. Then, we might argue, it doesn't matter that there is no specific default available for minefield crossings. Certainly it appears more psychologically plausible that we use general rather than specific defaults — the load on memory would be very great otherwise.[4]

Let us suppose for a moment that there is such a general "crossing" default — to the effect that if someone (or something) is crossing something, then the default is that they get to the other side of it. We might express this provisionally as:

"X crossing Y" > "X cross Y"

[4]This leaves us, however, needing to explain why we can't use the general "crossing" default for 'Mary was swimming the Atlantic'. We will consider this below.

where '>' signifies non-monotonic entailment.

What do we mean by saying that there "is" such a default? Does it exist out there in the world, independently of us? Is it part of the way we perceive things? Suppose one person "uses" a different set of defaults from someone else — for instance a community of religious believers might well have defaults regarding their religious beliefs that people outside that community generally don't share. A satisfactory account should enable us to make this precise, and we will attempt to do so shortly.

The notion of default[5] we are using looks rather like what has been called a **natural regularity** (see Barwise and Seligman 1994). A natural regularity is a constraint perceived by one or more agents as existing between kinds of things in the world — for example between the ringing of a doorbell and the fact that someone is standing on the porch, to take one of Barwise and Seligman (B&S)'s examples. We use such regularities all the time to allow us to reason with (often incomplete) information about the world we live in. An important point about natural regularities is that they are both **reliable** and **fallible**. In general, they work, which is why they are useful to us, but they "allow exceptions" and thus our reasoning may fail to reflect reality on a particular occasion.

Channel theory is a mathematical theory of information flow based on the notion of natural regularity, currently under development (see Barwise and Seligman 1994, Barwise 1993, Seligman and Barwise 1993). An important feature of channel theory is that it works on two levels — the level of tokens (particular events or objects, for example) and the level of types. This is the key to the treatment of exceptions — regularities exist at the level of types, but whether or not a connection between two tokens is of the appropriate type determines whether or not the associated reasoning goes through. For example, suppose there is a constraint or regularity between situations of a type we could call 'X crossing Y' (call this α) and those of type: 'X cross Y' (call this β). This is written:

$\alpha \Rightarrow \beta$ (using B&S (1994) notation).

Now consider a particular occasion where Mary is crossing the road. In general, we might expect her to get across — this is reflected by the above constraint.[6] Now consider the situation[7] (call it s) that corresponds to "Mary crossing the road" — it may support facts about her position at a given time, her direction of travel, her intentions, etc. s may be linked by a **signalling relation** to another situation s', which might be thought of

[5]We will argue below that "default" is not quite the right notion.
[6]We will consider the minefield question below.
[7]We employ the situation theoretic notion of situation here.

as a situation supporting facts about Mary which ends a few minutes later. This is denoted $s \to s'$, and is known as a **connection**.

Whether or not the particular connection is of type $\alpha \Rightarrow \beta$ is determined by a channel. A channel c contains a number of connections and a number of constraints, and tells us which connections are of which types. For example, the relevant channel c may tell us in this case that $s \to s' :_c \alpha \Rightarrow \beta$. If this is the case, then a soundness condition allows us to infer that $s' : \beta$. In other words, Mary gets to the other side. However, if c does not classify $s \to s'$ in this way, then we cannot conclude that $s' : \beta$, and this allows Mary not to reach the other side. This seems to capture very neatly the fact that progressives may sometimes "fail" in the sense that the complete event is not realised.

There may, however, be a problem in identifying the situation we have called s'. We proposed above that s' should be thought of as the situation supporting facts about Mary ending a few minutes after she is in the middle of the road. In other words, the "end time" of s' is associated with the time that Mary would have been expected to reach the other side, had she done so. If Mary does indeed reach the other side, then there is no problem in picking out the time of s'. However, as we explained above, the progressive sentence may still be true even if Mary never reaches the other side. It then becomes difficult to identify the end-time of s', because we have no way of knowing exactly when Mary would have reached the other side. This, in turn, makes it difficult to know exactly which situation we are calling s'.

In order to get around this problem, we propose that in cases where Mary does not reach the other side, we go a step further than saying merely that $s \to s'$ is not of type $\alpha \Rightarrow \beta$. We propose that in this case, s is a **pseudosignal** (in Barwise and Seligman's terminology) for the relevant constraint in channel c. This means that s is not connected to any other token in this channel — there simply is no s' such that $s \to s'$ in c. We can now express the semantics of the progressive as follows:

An utterance of a sentence of the form
'*X was V-ing Y*' can be felicitously used to describe a situation s iff:

There is a channel c and a type α such that $s : \alpha$,
and $\alpha \Rightarrow \beta$ is one of the types of c
(where β is the type of the corresponding "complete event", '*X V-ed Y*').

If there is a situation s' such that $s \to s'$ is one of the tokens of c, and if c classifies $s \to s'$ as of type $\alpha \Rightarrow \beta$, then by the soundness condition we can conclude that $s' : \beta$, which means that the complete event is realised.

However, as we saw above, $s \to s'$ need not be of type $\alpha \Rightarrow \beta$ in order for the progressive to be true — indeed, there may not even be an s'.

Thus an utterance of a progressive sentence may correctly describe a situation s even in cases where the complete event does not take place — which is exactly what we require.

Now let us look at some examples. We discussed above how these notions allow us to capture the semantics of:

(15) Mary was crossing the road.

To summarise briefly, we say that there is a channel c which contains a number of connections between situations. c also contains the constraint $\alpha \Rightarrow \beta$, where α is a situation type corresponding to an incomplete crossing, and β a situation type corresponding to a completed crossing. We proposed that if the described situation s is of type α, then s may be described using the appropriate progressive ('X was crossing Y').

Now we have to explain why this does not work for:

(16) Mary was swimming the Atlantic.

which is judged by speakers to sound odd[8] when used to describe a situation where, for example, Mary sets out from the coast of Cornwall and swims for several hundred yards in the direction of America. Why can't we use a "swimming constraint" that holds between "incomplete swims" and "complete swims" in the same way as for our crossing example above? The existence of such a swimming constraint appears intuitively as plausible as that of a crossing constraint — the constraints say nothing about what is being swum or crossed, remember. The point is that if we employ constraints at this level of generality, we have no way of ruling out the application of such a constraint to examples like (16). We want to keep the general swimming constraint (abbreviated to 'gsc'), but to rule out a less general instance of that constraint where the object is the Atlantic. Note that this is not the same as saying that a particular connection between situations is not classified as of the type of the constraint. This would only tell us that the complete event was not realised. It would not rule out the use of the progressive in the first place.

We propose the following explanation. Let us suppose that there exist (in the sense that human beings perceive them and use them to reason with) constraints at a fairly general level, that we might informally call, as above, crossing constraints, swimming constraints, etc. We do not want to commit ourselves at this stage to saying exactly what constraints we use, or at what level of generality we perceive them. Much further work would be required in order to establish this. However, we will make a number of tentative suggestions here in order to try to explain the data.

[8]Unless we know, for example, that Mary has superhuman capabilities.

Imagine someone hearing a progressive like (16). In order to make sense of it, the hearer starts with the general swimming constraint (gsc), which we might call $\gamma \Rightarrow \delta$, δ being the situation type we could label 'X swims Y', with X and Y unspecified. Now the hearer attempts to instantiate the gsc by substituting 'Mary' for X and 'the Atlantic' for Y. If successful, she will end up with a more specific constraint, $\gamma' \Rightarrow \delta'$, where δ' is the situation type we could label 'Mary swims the Atlantic'.

We propose that this last step is the stumbling block in the case of (16). The hearer is unable to instantiate the general constraint to obtain a more specific constraint in this case. The reason for this is that the hearer knows that the situation type 'Mary swims the Atlantic' is simply not realisable. This contrasts with:

(17) Mary was crossing the road.

where the hearer may readily instantiate the general crossing constraint (gcc) to give a specific constraint involving Mary crossing roads. The reason is that 'Mary crosses the road" is a seen as a realisable situation type — the hearer can conceive of it happening.

The impossibility of the instantiation in (16) is responsible for the fact that this is judged as an unacceptable progressive. Of course, if the hearer believes that Mary is superhuman, or receives divine assistance, then the instantiation may become possible, thus allowing her to accept the progressive. This allows us to account for the so-called 'miracle scenarios' discussed by Landman.

Now, what about Mary crossing her minefield? Can we explain why:

(18) Mary was crossing the minefield.

is acceptable? Once again, the hearer starts with the gcc, $\alpha \Rightarrow \beta$. Now she tries to instantiate X and Y to 'Mary' and 'the minefield' respectively. The criterion for doing this is that 'Mary cross the minefield' is a realisable situation type. We do not require it to be the default outcome, nor that it is even likely or expected, but simply that it is (in the mind of the hearer) a **possible** outcome. Provided that the hearer believes it is possible for Mary to get to the other side of the minefield, the more specific constraint may be derived, and the progressive is licensed. Even though the hearer might think the most likely outcome is for Mary to get blown up, provided she believes it is **possible** for Mary to get across, the progressive is judged acceptable.

This would explain why most speakers are happy to accept (18), while rejecting (16). We differ from Asher in that we do not require there to be a default that Mary gets to the other side of the minefield. We also differ from Landman in that we do not require Mary's getting to the other side to be a "reasonable outcome", but simply possible.[9]

It should be noted that if we force ourselves to interpret '$\Rightarrow$' as a default entailment, then instantiating the gcc gives:

'Mary crossing the minefield' $\Rightarrow$ 'Mary crossed the minefield'

so we are back with a default account like Asher's, which we have argued against. Thus it seems that we should not force a default interpretation on '$\Rightarrow$'. Perhaps we can see general constraints like the gcc, but not the more specific constraints obtained by instantiation, as defaults. For example, it seems reasonable to view the gcc as a default, but not the more specific constraint obtained by substituting the object variable Y with 'a minefield'. If we were then to go on to substitute the subject position X with 'the army captain', knowing that this individual has mine-detecting equipment and expertise, then this even more specific constraint might once again be seen as a default. What appears important is that we do not necessarily interpret '$\Rightarrow$' as a default.

Another point we should note is that it seems wrong to think of the hearer's interpretation of the progressive as necessarily involving default reasoning, even in cases where a default appears appropriate. It is perfectly possible to hear a progressive like (17) and not infer that Mary got to the other side, even if the speaker did not give information to the contrary. Indeed, one reason for the speaker choosing a progressive may be to indicate that the event was not completed, especially given that the progressive is the marked form in English.[10]

We can now give a revised semantics for the progressive, taking into account the notion of instantiation of general constraints discussed above.

An utterance of a sentence *sent* of the form
'*X was V-ing Y*' can be felicitously used to describe a situation s iff:

There is a channel c and a type α such that $s : \alpha$, and $\alpha \Rightarrow \beta$ is one of the types of c,
(where β is the type of the corresponding "complete event", '*X V-ed*

[9] It could probably be argued that our account is not so very different from Landman's in this respect — that the notions of "possible" and "reasonable" are not so far apart. This may well be true in general, although we have argued that Landman's account would not predict the acceptability of the minefield example. But a more important point is that, although there are some strong intuitive similarities between Landman's account and ours, the channel theoretic framework gives us tools to begin to make precise some notions which are left unformalised in Landman's account.

[10] Of course, there are other reasons for choosing the progressive, too, such as to convey temporal inclusion or what we call 'backgrounding' in Glasbey 1994.

Y), and:
$\alpha \Rightarrow \beta$ can be instantiated to $\alpha' \Rightarrow \beta'$ by substituting the values of X and Y in *sent*,
where $s : \alpha'$ and β' is a "realisable" or "conceivably possible" situation type.

Perspectives

Earlier, we discussed examples like:

(19) We were flying to Manchester / We were flying to Havana

We will now see how the notion of channel allows us to deal with such examples. We saw above that a channel c contains a number of connections, a number of constraints, and determines which connections are of which types. Thus a channel may be thought of as a perspective — one particular way of classifying reality.

Suppose we say that in one channel, c_1, there's a constraint between what we were doing (sitting in the plane, with the appropriate tickets) and flying to Manchester. In another, c_2, there's a constraint between what we were doing (sitting in the plane, with the appropriate tickets, and hijackers on board with destination Havana) and flying to Havana. Thus each "perspective" in the informal sense used by Landman corresponds to a distinct channel. The different channels correspond to alternative perspectives that may be taken by the speaker depending on her knowledge or point of view at the time of evaluation.

Both perspectives are acceptable in this case, each corresponding to a different channel. Of course, both outcomes cannot be realised as they are incompatible. c_2 is relevant with hindsight — when we know the hijackers were on board. Of course, many questions remain about exactly how we reason with channels. The answer to some of these may have to wait for further developments in channel theory. But here, at least, we have a promising start to the analysis of problematic progressives of this type.

16.3 More Examples

In this section we will look briefly at some further examples which are problematic for Landman, Asher or both, and sketch out how the channel theoretic account helps us address some of the problems.

Let us first see how the channels proposal would work in the case of examples where **intentions** are important, such as:

(20) John was walking to the shops.

In Section 1 we described a scenario where this progressive may not felicitously be used, because John's intentions are "wrong". First, let us suppose that there is a general "walking-to" constraint $\phi \Rightarrow \psi$, where ψ is the type of situation 'X walks to Y'. In order to capture the requirement that John's intentions must be right, we can build information about the intentions of X into ϕ. Thus a situation s will only be of type ϕ if s supports the information that X intends to walk to Y. This prevents a situation where John intends to walk to the park from being classified as of type ϕ, and hence the progressive cannot be licensed.

In contrast, we have examples like:

(21) Mary was making John a millionaire.

where the progressive may be true even when the agent does not intend the outcome to happen. In this case, suppose we have a general constraint $\phi' \Rightarrow \psi'$, where ψ' is 'X makes Y a millionaire'. In this case we do not need to place any restrictions on ψ' concerning the intentions of X.[11]

Now let us look at an example discussed by Lascarides (1991):

(22) Max was winning the race.

This is an interesting progressive because, as Lascarides points out, it is possible to use:

(23) Max was not winning the race.

to refer to a time t during the race, when in fact Max goes on to win the race at a later time. Lascarides observes that this is problematic for accounts of the progressive based on the notion of "eventual outcome", among which Landman's account is numbered.[12] The problem for Landman is that if Max wins the race in the actual world, we should be able to predicate 'Max was winning the race' of any time during the race. However, this is clearly not the case. We can say:

(24) Max was not winning the race at the halfway point, but he put on a sprint finish and won in the end.

We might note that 'win the race' is a rather unusual achievement in this respect and others.[13]

An initial survey suggests that there are not many achievements that pattern like 'win the race' in this respect. 'Be winning the race (at t)' seems to mean something very close to 'be in a good position to win at t'

[11] Clearly, the constraint used here must be sufficiently specific to allow discrimination between cases where intention must be present, and cases where it need not. This is why we have suggested that the constraint specifies 'millionaire' rather than 'X makes Y a Z'.

[12] Vlach (1981) discusses such examples.

[13] For example, we can say:

(25) John was winning the race for the first few minutes.

Compare, however:

(26) *??John was reaching the summit for the first few minutes (of his climb).*

or 'be in the lead at t' — while somehow acknowledging that there may not be a very strong connection between being in a good position to win and winning — something every sports enthusiast knows.

Does our channels account provide a way of capturing this? An intitial attempt at an explanation might go as follows. Suppose there is a "winning constraint" which holds between the situation type 'X is in the lead' and the situation type 'X wins'. The existence of such a constraint reflects the fact that we perceive there to be a connection (albeit not necessarily a very strong one) between situations of being in the lead and situations of winning. The existence of such a constraint licenses the use of 'John was winning the race' to describe a situation of being in the lead. However, the fact that it is only perceived to be a weak constraint means that there will be many cases where it does not hold, and where being in a winning position at t does not signal winning the race.

This is only a preliminary analysis and more work needs to be done on progressives like 'win the race'. However, it is encouraging that the use of constraints (which need not necessarily be seen as defaults, but may be much "weaker" connections) has allowed us to make some progress here.

Finally, let us look briefly at some examples that Ogihara (1990) identified as problematic for analyses of the progressive. These include:

(27) Mary is marking 50 exam papers.

which is an acceptable progressive, usable to describe a situation where Mary has marked, say, eleven papers and is taking lunch before starting work on the twelfth.

Compare, however:

(28) ??Mary is drinking 3 cups of tea and 5 glasses of wine.

which sounds distinctly odd if I say it in the middle of a day when Mary does in fact consume all these drinks — unless I can somehow see it as part of a plan of Mary's to do so.[14] Suppose someone has bet Mary a tenner that she couldn't possibly achieve this feat, and she is in the middle of attempting it when she begins to feel sick. Then she might say afterwards:

(29) I was drinking 3 cups of tea and 5 glasses of wine when I started to feel sick and gave up.

In this context, the progressive sounds perfectly natural. In a comment on Ogihara's paper, Caenepeel and Moens (1990) suggest that an account of the progressive needs to capture the difference between something that

[14]Ogihara suggests that progressives like these are acceptable if we can regard the eating and drinking as comprising one "coherent event". He does not make this notion precise, however.

can be seen as part of a plan[15] and something that can't. They take this to suggest that the semantics of the progressive cannot be captured in purely temporal terms, but they do not formalise the notion of 'plan'.

Can we make some progress here using channels? One way of thinking about it is to say that, in the case of a pre-planned scenario like the marking of a set of papers, a constraint is perceived between the type of situation where one is in the middle of such a task (e.g., halfway through the twelfth paper) and the type of situation that corresponds to completing the task. Similarly, if we know that Mary is attempting some feat of beverage consumption, then we can perceive a constraint between the type of situation where she is partway through the drinking and the situation type where she completes the task. If, however, the drinking was not planned but simply the result of Mary having a bad day[16], then we can perceive no such constraint and the progressive is not licensed.

Notice that the notion of default does not seem to work here. A hearer may accept (28) while being very sceptical of Mary's ability to achieve the feat. Thus we have another case where it seems we do not want to identify the perception of constraints with default reasoning.

16.4 Conclusion

We have reviewed recent accounts of the progressive and argued that they are unsatisfactory in certain respects. In attempting to decide what was going wrong, we showed how moving from a default analysis like Asher's to an account based on the notion of natural regularity allows us to deal with problems for Asher's analysis while at the same time capturing some important insights of both Asher's and Landman's approaches. We showed, too, how channel theory can make precise the notion of perspective introduced informally by Landman. A range of examples has been examined here and analyses sketched within a channel theoretic framework. Further work will involve applying the channel theoretic account to wider range of examples.

References

Asher, Nick. 1992. A default, truth conditional semantics for the progressive. *Linguistics and Philosophy* 15:463–508.

Barwise, Jon. 1993. Constraints, channels and the flow of information. In *Situation Theory and Its Applications, 3*, ed. P. Aczel, D. Israel, Y. Katagiri, and S. Peters. 3–27. CSLI Lecture Notes. Stanford: CSLI Publications.

Barwise, Jon, and Jerry Seligman. 1994. The rights and wrongs of natural regularity. In *Philosophical Perspectives*, ed. J. Tomberlin.

[15]However, they point out that the problem of speaking of a plan is that it has connotations of intentionality.

[16]Or, perhaps, a good party ...

Bennett, J., and B. Partee. 1972. Toward the logic of tense and aspect in English. Indiana University Linguistics Club, Bloomington, Indiana.

Caenepeel, Mimo, and Marc Moens. 1990. Progressives, perfects and the temporal structure of discourse. In *Tense and Aspect in English*, ed. Hans Kamp. 39–47. Centre for Cognitive Science, University of Edinburgh: Dyana Report R1.2.A.

Cooper, Robin. 1985. Aspectual classes in situation semantics. Technical Report 84-14C. CSLI, March.

Dowty, David. 1979. *Word Meaning and Montague Grammar.* D.Reidel, Dordrecht.

Glasbey, Sheila. 1994. Progressives, Events and States. In *Proceedings of the Ninth Amsterdam Colloquium*, ed. P. Dekker and M. Stokhof, 313–332. Department of Philosophy, University of Amsterdam. Institute for Logic, Language and Computation.

Hinrichs, Erhard. 1983. The semantics of the English progressive: a study in situation semantics. In *Papers from the Nineteenth Regional Meeting of the Chicago Linguistics Society*, ed. A. Chukerman, M. Marks, and J. F. Richardson, 171–182.

Landman, Fred. 1992. The progressive. *Natural Language Semantics* 1:1–32.

Lascarides, Alex. 1991. The progressive and the imperfective paradox. *Synthese* 87(3):401–407.

Ogihara, Toshiyuki. 1990. The semantics of the progressive and the perfect in English. In *Tense and Aspect in English*, ed. Hans Kamp. 3–38. Centre for Cognitive Science, University of Edinburgh: Dyana Report R2.3A.

Seligman, Jerry, and Jon Barwise. 1993. Channel theory: towards a mathematics of imperfect information flow. Unpublished ms.

Vendler, Zeno. 1967. Verbs and times. In *Linguistics in Philosophy.* Chap. 4, 97–121. Cornell University Press, Ithaca, NY.

Vlach, F. 1981. The semantics of the progressive. In *Syntax and Semantics, Vol.14:Tense and Aspect*, ed. P. J. Tedeschi and A. E. Zaenen. New York: Academic Press.

17

This Might Be It

JEROEN GROENENDIJK, MARTIN STOKHOF AND FRANK VELTMAN

Introduction

Discussions often end before the issues that started them have been resolved. For example, in the late sixties and early seventies, a hot topic in philosophical logic was the development of an adequate semantics for the language of modal predicate logic. However, the result of this discussion was not one single system that met with general agreement, but a collection of alternative systems, each defended most ably by its proponents

Although it would seem that the topic has lost much of its controversial status, this paper adds one more system to the existing stock. It offers a semantics for the language of modal predicate logic, which is new, not in the sense that it proposes a new ontology as an alternative to the possible world paradigm, but new because it characterizes the meaning of a sentence in terms of its information change potential rather than its truth conditions. What we hope to show is that this dynamic twist sheds new light on old issues concerning modality, (co)reference, identity and identification.[1]

We owe a special thanks to Paul Dekker The present paper builds heavily on the last chapter of his thesis His comments on various stages of the work reported here have prevented us from making many mistakes For the remaining ones we take the blame Maria Aloni and Jelle Gerbrandy also provided useful feed-back Earlier versions of the paper were presented on various occasions The first of these was the Workshop on Tense and Modality (Columbus, Ohio, July 1993) For their helpful comments, we thank the participants of that workshop, and of other events where we talked about this material The work reported here was supported by ESPRIT Basic Research Action 6852, DYANA-2 A more elaborate version of the paper appears in DYANA-deliverable R2 1B, Fall 1994

[1]The present system combines dynamic predicate logic with update semantics for modal propositional logic (see Groenendijk and Stokhof 1991, Veltman 1990)

Logic, Language and Computation
Jerry Seligmann and Dag Westerståhl, eds

The idea that meaning is information change is implemented by interpreting sentences as updates, as functions from information states to information states. From this epistemic perspective, the notion of truth, which relates language to the world, loses its key role. Central notions are consistency and support, which relate language, not to the world, but to the information language users have about it. Consequently, the entailment relation has to face change, too.

Since interpretation is viewed as a process of updating information, the structure and contents of information states has to be explicated. This is done section 1. In section 2, an update semantics for the language of modal predicate logic is stated, which is illustrated by a discussion of some representative examples in section 3. In section 4, special attention is paid to problems of identity and identification, some of which, it is argued, necessitate the introduction of demonstratives. In section 5 we look ahead.

17.1 Information

An information state is looked upon as a set of possibilities, viz., those alternatives which are still open according to someone in that state. What the possibilities are depends on what the information is about. First, there is information about the world. Quite often, such information is gathered by verbal means. The interpretation of discourse raises its own questions. For example, there is the issue of resolving anaphoric relations. We have to keep track of what we talk about. This kind of discourse information is more like a book-keeping device than like real information. Yet, it is essential for the interpretation of discourse, and since the latter is an important source of information about the world, discourse information, indirectly, also provides such information.

Information about the world is represented as a set of possible worlds, those worlds that given the information available still might be the real one. Worlds are identified with first order models. In the present paper, it is assumed that the language users know which objects constitute the domain of discourse (although they may not know their names). In view of this, all possible worlds share one domain. Hence, a possible world can be identified with an interpretation function of a first order model. Extending information about the world amounts to eliminating worlds from the ones which were still considered possible.

The language interpreted is a logical language with quantifiers and variables. The use of a quantifier introduces a new item of conversation, a new *peg*. Variables are the anaphoric expressions of the language. To enable the

On the problem of combining the two see, e.g., Groenendijk and Stokhof 1990, Dekker 1993, chapter 5, Vermeulen 1994, chapter 2.

resolution of anaphoric relations, discourse information keeps track, not only of the number of pegs, but also of the association between variables and pegs. Extending discourse information is adding variables and pegs.

Discourse information is linked to information about the world, *via* possible assignments of objects to the pegs (and hence, indirectly, to the variables associated with these pegs). In general, not every assignment of an object to a peg is possible — both the discourse and the available information may provide restrictions —, but usually, more than one is. Getting better informed on this score is eliminating possible assignments. Suppose a certain assignment is the only one left with respect to some world which is still considered possible. In that case elimination of the assignment brings along the elimination of the world. This is how discourse information may provide information about the world.

17.1.1 Information States

In the possibilities that make up an information state, the discourse information is encoded in a *referent system*,[2] which tells which variables are in use, and with which pegs they are associated. We use natural numbers as pegs.

Definition 1 A *referent system* is a function r, which has as its domain a finite set of variables v, and as its range a number of pegs.
If the number of pegs in a referent system is n, then the numbers $m < n$ are its pegs.

The use of a quantifier $\exists x$ introduces the next peg, and associates the variable x with that peg:

Definition 2 Let r be a referent system with domain v and range n.
$r[x/n]$ is the referent system r' which is like r, except that its domain is $v \cup \{x\}$, its range is $n+1$, and $r'(x) = n$.

Note that it is not excluded that x is already present in v. This situation occurs if the quantifier $\exists x$ has been used before. In that case, even though the variable x was already in use, it will be associated with a new peg. The peg that x was connected with before remains, but is no longer associated with a variable. This means that a referent system r is an injection.

Associating a variable with a new peg is the proto-typical way in which a referent system is extended:

Definition 3 Let r and r' be two referent systems with domain v and v', and range n and n', respectively.

---

[2]The use of referent systems is inspired by the work of Kees Vermeulen. See Vermeulen to appear, Vermeulen 1994, chapter 3.

r' is an *extension* of r, $r \leq r'$, iff $v \subseteq v'$; $n \leq n'$; if $x \in v$ then $r(x) = r'(x)$ or $n \leq r'(x)$; if $x \notin v$ and $x \in v'$ then $n \leq r'(x)$.

A referent system $r[x/n]$ is always a *real* extension of r.

Above, a distinction was made between discourse information, information about the world, and a link between the two. These three ingredients are present in the *possibilities*, which in turn make up information states.

Definition 4 Let D, the *domain of discourse*, and W, the set of *possible worlds*, be two disjoint non-empty sets.
The *possibilities* based on D and W is the set I of triples $\langle r, g, w\rangle$, where r is a referent system; g is a function from the range of r into D; $w \in W$.

The function g assigns an object to each peg in the referent system. The composition of g and r indirectly assigns values to the variables that are active: $g(r(x)) \in D$.

Information states are (real) subsets of the set of possibilities:

Definition 5 Let I be the set of possibilities based on D and W.
The set of *information states* based on I is the set S such that $s \in S$ iff $s \subseteq I$, and $\forall i, i' \in s\colon i$ and i' have the same referent system.

Variables and pegs are introduced globally with respect to information states. That is why an information state has a unique referent system.

17.1.2 Information Growth

One way in which information can grow is by adding variables and pegs and assigning some object to them:

Definition 6 Let $i = \langle r, g, w\rangle \in I$; n the range of r; $d \in D$, $s \in S$.

1. $i[x/d] = \langle r[x/n], g[n/d], w\rangle$
2. $s[x/d] = \{i[x/d] \mid i \in s\}$

Information can grow, not just by adding discourse information, but also by eliminating possible assignments of objects to pegs, and by eliminating possible worlds.

Definition 7 Let $i, i' \in I, i = \langle r, g, w\rangle$ and $i' = \langle r', g', w'\rangle$, and $s, s' \in S$.

1. i' is an *extension* of i, $i \leq i'$ iff $r \leq r'$, $g \subseteq g'$, and $w = w'$
2. s' is an *extension of* s, $s \leq s'$ iff $\forall i' \in s'\colon \exists i \in s\colon i \leq i'$

The extension relation is a partial order. There is a unique minimal information state, the *state of ignorance* in which all worlds are still possible and no discourse information is available yet: $\mathbf{0} = \{\langle \emptyset, \emptyset, w\rangle \mid w \in W\}$. The maximal element in the ordering $\mathbf{1} = \emptyset$, the *absurd state* in which no possibility is left. Less maximal, but more fortunate, are states of *total information*, consisting of just one possibility.

Some auxiliary notions:

Definition 8 Let $s, s' \in S, s \leq s', i \in s, i' \in s'$.

1. If $i \leq i'$, we say that i' *is a descendant of* i *in* s'
2. If i has one or more descendants in s', we say that i *subsists in* s'
3. If all $i \in s$ subsist in s', we say that s *subsists in* s'

17.2 Updating Information States

Now that information states are defined, they can be put to use: in updating. The formulae of the familiar language of modal predicate logic are interpreted as (partial) functions from information states to information states.

We use postfix notation: $s[\phi]$ is the result of updating s with ϕ, $s[\phi][\psi]$ is the result of first updating s with ϕ, and next updating $s[\phi]$ with ψ. Whether s can be updated with ϕ may depend on the fulfillment of certain constraints. If a state s does not meet them, then $s[\phi]$ does not exist, and the interpretation process comes to a halt.

The possibilities contain all that is needed for the interpretation of the basic expressions of the language: individual constants, variables, and n-place predicates.

Definition 9 Let α be a basic expression, $i = \langle r, g, w\rangle \in I$, with v the domain of r, and I based upon W and D.

1. If α is an individual constant, then $i(\alpha) = w(\alpha) \in D$
2. If α is a variable such that $\alpha \in v$, then $i(\alpha) = g(r(\alpha)) \in D$, else $i(\alpha)$ is not defined
3. If α is an n-place predicate, then $i(\alpha) = w(\alpha) \subseteq D^n$

The absence of a variable in the referent system of a state will be the only source of partiality of updates.

The following definition specifies the update semantics for the language of modal predicate logic:

Definition 10

1. $s[Rt_1 \ldots t_n] = \{i \in s \mid \langle i(t_1), \ldots, i(t_n)\rangle \in i(R)\}$
2. $s[t_1 = t_2] = \{i \in s \mid i(t_1) = i(t_2)\}$
3. $s[\neg\phi] = \{i \in s \mid i$ does not subsist in $s[\phi]\}$
4. $s[\phi \wedge \psi] = s[\phi][\psi]$
5. $s[\exists x\phi] = \bigcup_{d\in D}(s[x/d][\phi])$
6. $s[\Diamond\phi] = \{i \in s \mid s[\phi] \neq \emptyset\}$

In updating an information state with an atomic formula, those possibilities are eliminated in which the objects denoted by the arguments do not stand in the relation expressed by the predicate. The same holds for identity statements: those possibilities are eliminated in which the two terms do not denote the same object.

Notice that atomic updates can be partial. If one of the argument terms is a variable that is not present in the referent system of the information state, then its denotation is not defined, and hence the update does not exist. This carries over to all the other update clauses. If somewhere in the interpretation process we meet a variable that at that point has not been introduced, then the whole process comes to a halt.

In calculating the effect of updating a state s with $\neg\phi$, s is updated hypothetically with ϕ. Those possibilities that subsist after this hypothetical update are eliminated from the original state s.

Updating a state with a conjunction is a sequential operation: the state is updated with the first conjunct, and next the result of that is updated with the second conjunct. The interpretation of a conjunction is the composition of the update functions associated with the conjuncts.

If a state s is updated with $\exists x\phi$, its referent system is extended with a new peg, and the variable x is associated with that peg. An object d is selected from the domain and assigned to the newly introduced peg. The state $s[x/d]$ that results from this is updated with ϕ. After this is done for every object d, the results are collected.

The operator $\Diamond$ corresponds to the epistemic modality *might*. Updating a state s with $\Diamond\phi$, amounts to testing whether s can be consistently updated with ϕ. If the test succeeds, the resulting state is s again; if the test fails because updating s with ϕ results in the absurd state, then s updated with $\Diamond\phi$ results in the absurd state.

The semantics just presented defines the interpretation of the formulae of the language in terms of their information change potential. Actually, they change information states in a particular way:

Observation 1 *For every formula ϕ and information state s: $s \leq s[\phi]$*

The observation tells that we are justified in calling the semantics an *update semantics*. The interpretation process always leads to an information state that is an extension of the initial state.

Other logical constants can be added by definition in the usual way. Calculating the definitions out, we get:

Observation 2

1. $s[\phi \rightarrow \psi] = \{i \in s \mid$ *if i subsists in $s[\phi]$, then all descendants of i in $s[\phi]$ subsist in $s[\phi][\psi]$*$\}$
2. $s[\phi \vee \psi] = \{i \in s \mid$ *i subsists in $s[\phi]$ or i subsists in $s[\neg\phi][\psi]$*$\}$
3. $s[\forall x\phi] = \{i \in s \mid$ *for all $d \in D$: i subsists in $s[x/d][\phi]$*$\}$
4. $s[\Box\phi] = \{i \in s \mid$ *s subsist in $s[\phi]$*$\}$

It should be remarked that it is not possible to make a different choice between basic and defined operations that leads to the same overall results.

17.2.1 Consistency, Support, and Entailment

Truth and falsity concern the relation between language and the world. In update semantics it is information about the world rather than the world itself that language is related to. Hence, the notions of truth and falsity cannot be expected to occupy the same central position as they do in standard semantics. More suited to the information oriented approach are the notions of consistency and support.

For a hearer to be willing to update with a sentence, the update should not lead to the absurd state. And if a speaker is to assert a sentence correctly, it should not constitute a 'real' update in her information state.

Definition 11 Let s be an information state.

1. s *allows* ϕ iff $s[\phi]$ exists and $s[\phi] \neq \emptyset$
2. s *supports* ϕ iff $s[\phi]$ exists and s subsists in $s[\phi]$
3. s *forbids* ϕ iff $s[\phi] = \emptyset$

With respect to the rare states of total information about the world, the notions of being allowed and supported coincide, and could be equated with truth, for non-modal statements that is. Likewise, in such states being forbidden amounts to falsity.

According to semantic intuition, a discourse is unacceptable if there is not at least *some* state that allows it. And if a sentence is not supported by *any* non-absurd state, which means that no speaker could sincerely utter it, then that sentence is judged unacceptable.

Definition 12

1. ϕ is *consistent* iff there is some information state which allows ϕ
2. ϕ is *coherent* iff there is some non-absurd state that supports ϕ

Note that coherence implies consistency. Concerning the acceptability of a single sentence, it would suffice to require coherence. Still, it makes sense to distinguish both notions. A discourse may consist of a sequence of sentences, possibly uttered by different speakers in different information states. The acceptability of a discourse minimally requires that one by one the sentences are coherent. That does not imply that the discourse as a whole can be supported by a single information state. Hence, it does not imply that the discourse as a whole is consistent. The latter is an independent requirement for the acceptability of a discourse.

Since discourse consistency and sentence coherence are necessary conditions for acceptability, these semantic properties present us with criteria for testing the adequacy of a proposed semantics.

For this purpose, the notion of entailment is just as important. Entailment is not defined in the usual way in terms of truth, but in terms of sequential update and support:

Definition 13 $\phi_1, \ldots, \phi_n \models \psi$ iff for all information states s such that $s[\phi_1] \ldots [\phi_n][\psi]$ exists, it holds that $s[\phi_1] \ldots [\phi_n]$ supports ψ

Below, some of the properties of the entailment relation are illustrated.

17.2.2 Equivalence

A suitable notion of equivalence may be expected to tell when two expressions can be substituted for each other in a meaning preserving way. Within update semantics, meaning is preserved if the update effects are. This being so, the usual definition of equivalence in terms of mutual entailment cannot be used. For example, $\exists xPx$ and $\exists yPy$ mutually entail each other, and so do $\exists xPx$ and Px, but, obviously, they cannot be replaced for each other in all contexts preserving update effects.

At the same time, it will also not do to require that ϕ and ψ are equivalent if they have exactly the same update effects. Under such a definition, $\exists x \exists y Rxy$ and $\exists y \exists x Rxy$ would not come out equivalent, and neither would $\exists xPx$ and $\exists xPx \wedge \exists xPx$. The reason for this is that the referent system of an information state not just keeps track of which variables and pegs are present, but also of the order in which they were introduced. Furthermore, there can be pegs around that are no longer associated with a variable. In view of this, the resulting information states are not required to be the *same*, but to be *similar*, where the notion of similarity ignores these less important differences between information states.

Definition 14 Let $i, i' \in I, i = \langle r, g, w \rangle, i' = \langle r', g', w' \rangle$, with v and v' the domain of r and r', respectively; and let $s, s' \in S$.

1. *i is similar to i'* iff $v = v'$, $w = w'$, and $\forall x \in v$: $g(r(x)) = g'(r'(x))$
2. *s is similar to s'* iff $\forall i \in s$: $\exists i' \in s'$: i is similar to i' and $\forall i' \in s'$: $\exists i \in s$: i' is similar to i

Similarity is an equivalence relation.

Definition 15 $\phi \equiv \psi$ iff for all information states s: $s[\phi]$ is similar to $s[\psi]$.

17.3 Illustrations

A characteristic feature of update semantics, is that it can account for the fact that order matters in discourse. Consider:

(5) It might be raining outside [...] It isn't raining outside.

(6) It isn't raining outside [...] *It might be raining outside.

Given the sequential interpretation of conjunction, and the interpretation of the *might*-operator as a consistency test, the unacceptability of (2) is readily explained. After an information state is updated with the information that it is not raining, it is no longer consistent with our information that it might be raining. If, as in (1), things are presented in the opposite order, there is no problem.

So, the difference between (1) and (2) is explained by the following fact:

Observation 3 *Whereas* $\Diamond p \wedge \neg p$ *is consistent,* $\neg p \wedge \Diamond p$ *is inconsistent.*

Note that the dots in example (1) are important. If they are left out, or replaced by 'and', one is more or less forced to look upon (1) as a single utterance, of a single speaker, on a single occasion. But in that case, (1) intuitively is no longer acceptable. The following fact explains this:

Observation 4 *Although consistent,* $\Diamond p \wedge \neg p$ *is incoherent.*

An utterance of a sentence is incoherent if no *single* information state can support it.

Another way to look at the consistency and incoherence of $\Diamond p \wedge \neg p$ is as follows. Since it is consistent, there are states that can be updated with it. But once updated, such states cannot confirm what was said. For any non-absurd state s, $s[\Diamond p \wedge \neg p]$ does not support $\Diamond p \wedge \neg p$. This means that $\Diamond p \wedge \neg p$ is not idempotent:

Observation 5 $\Diamond p \wedge \neg p \not\models \Diamond p \wedge \neg p$

The reason behind all this is the *non-persistence* of formulae of the form $\Diamond\phi$: A state s may support $\Diamond\phi$, whereas a more informative state s' may be inconsistent with it. The non-persistence of modal formulae causes non-monotonicity of entailment:

Observation 6 *Although* $\Diamond p \models \Diamond p$*, we have that* $\Diamond p, \neg p \not\models \Diamond p$

Commutativity, idempotency and monotonicity, also fail to hold for reasons having to do with coreference rather than modality. For example, whereas $\neg Px \wedge \exists x Px$ is consistent, $\exists x Px \wedge \neg Px$ is not. And notice that $\neg Px \wedge \exists x Px$ is not idempotent. Finally, although $\exists x Px \models Px$, we have that $\exists x Px, \exists x \neg Px \not\models Px$.

17.3.1 Coreference and Modality

It is a characteristic feature of dynamic semantics that existential quantifiers can bind variables outside their scope. The variable in the second conjunct of (3) is bound by the quantifier in the first conjunct:

(7) $\exists x Px \wedge Qx$

Let n be the number of pegs in an information state s. First, s is updated with $\exists x Px$. Each possibility $\langle r, g, w\rangle \in s$ will have as many possibilities $\langle r[x/n], g[n/d], w\rangle$ as its descendants in $s[\exists x Px]$ as there are objects $d \in D$

such that $d \in w(P)$. From those, the update with Qx eliminates the ones in which $d \notin w(Q)$.

Exactly the same happens when s is updated with $\exists x(Px \wedge Qx)$:

Observation 7 $\exists x Px \wedge Qx \equiv \exists x(Px \wedge Qx)$

With the aid of the extended binding power of the existential quantifier a compositional and incremental account of cross-sentential anaphora can be given, and the same holds for donkey-anaphora.

Observation 8 $\exists x Px \to Qx \equiv \forall x(Px \to Qx)$

Equivalences such as these are characteristic of dynamic predicate logic.

Modal operators are transparent to the extended binding force of existential quantifiers. In (4), the occurrence of the variable within the scope of the *might*-operator is bound by the quantifier in the first conjunct:

(8) $\exists x Px \wedge \Diamond Qx$

In this case, the second conjunct only tests whether in the state that results after updating with the first conjunct there is at least one possibility $\langle r[x/n], g[x/d], w\rangle$ such that $d \in w(Q)$. In particular, this means that among the possible values of x after updating with the whole sequence, there may be objects d that in no w have the property Q.

As is to be expected, both (5) and (6) are inconsistent:

(9) $\exists x Px \wedge \Diamond\neg Px$

(10) $\exists x Px \wedge \Diamond\forall y \neg Py$

But the following formula is *not* inconsistent:

(11) $\exists x Px \wedge \forall y \Diamond\neg Py$

Suppose the domain consists of just two objects, and that according to some information state just one of them has the property P, but that it does not decide which one. Then for each of these objects it holds that it might not have the property P.

However, unlike (7), (8) *is* inconsistent:

(12) $\exists x(Px \wedge \forall y \Diamond\neg Py)$

The brackets make a difference. In updating a state s with (8), some object d is chosen, and $s[x/d][Px \wedge \forall y \Diamond\neg Py]$ is performed. In all possibilities that remain after updating $s[x/d]$ with Px, d has the property P. But then $\forall y \Diamond\neg Py$ will be inconsistent with $s[x/d][Px]$. And this holds for each choice of d. Hence (8) is inconsistent.

This means that dynamic modal predicate logic lacks some features which characterize dynamic predicate logic. This point may be elaborated.

Imagine the following situation. You and your spouse have three sons. One of them broke a vase. Your spouse is very anxious to find out who did it. Both you and your spouse know that your eldest didn't do it, he was playing outside when it must have happened. Actually, you are not interested in the question who broke the vase. But you are looking for your eldest son to help you do the dishes. He might be hiding somewhere.

In search for the culprit, your spouse has gone upstairs. Suppose your spouse hears a noise coming from the closet. If it is the shuffling of feet, your spouse will know that someone is hiding in there, but will not be able to exclude any of your three sons. In that case your spouse could utter:

(13) There is someone hiding in the closet. He might be guilty.

$\exists x Qx \wedge \Diamond Px$

But the information state of your spouse would not support:

(14) There is someone hiding in the closet who might be guilty.

$\exists x (Qx \wedge \Diamond Px)$

However, if the noise is a high-pitched voice, things are different. Now, your spouse knows it can't be your eldest, he already has a frog in his throat. In that case your spouse *can* say (10).

This also means that if your spouse yells (10) from upstairs, you can stay were you are, but if it is (9), you might run upstairs to check whether it is perhaps your aid that is hiding there.

So, there is a difference between (9) and (10),[3] and the semantics accounts for it:

Observation 9 $\exists x Px \wedge \Diamond Qx \not\equiv \exists x (Px \wedge \Diamond Qx)$

A similar observation applies to the following pair of examples.

(15) If there is someone hiding in the closet, he might be guilty.

$\exists x Qx \rightarrow \Diamond Px$

(16) Anyone who is hiding in the closet might be guilty.

$\forall x (Qx \rightarrow \Diamond Px)$

Take the same situation again. Only in case your spouse heard some high-pitched voice, (12) is a correct utterance. In the other case, (12) is not supported by the information state of your spouse, and only (11) is left.

Observation 10 $\exists x Qx \rightarrow \Diamond Px \not\equiv \forall x (Qx \rightarrow \Diamond Px)$

These facts are significant for at least two reasons. First, unlike in the predicate logical fragment of the language, in the full language it makes a difference whether a bound variable is inside or outside the scope of the quantifier that binds it. Secondly, since in any static semantics a variable can only be bound by a quantifier if it is inside its scope, it can never account for such differences.

[3] And we thank David Beaver for pointing this out to us.

There are two features of the proposed semantics which together are responsible for this result. The first is that the consistency test performed by the *might*-operator not only checks whether after an update with the formula following the *might*-operator there will be any worlds left, but also whether there will be any assignments left. Thus, even in a situation in which knowledge of the world is complete, epistemic qualification of a statement may still make sense. Example:

(17) $\exists x(x^2 > 4) \wedge \Diamond(x > 2) \wedge \Diamond(x < -2)$

Consider the world that results when the the operations and relations mentioned in (13) are given their standard interpretation in the domain of real numbers. In that case (13) will be supported by any state consisting of possibilities in which only this world figures.

The second feature is that existential quantification is *not* interpreted in terms of *global* (re-)assignment. Global reassignment, which would give wrong results, reads as follows:

$$s[\exists x\phi] = (\cup_{d\in D} s[x/d])[\phi]$$

Updating with $\exists x\Diamond Px$ would output *every* $d \in D$ as a possible value for x, as long as there is *some* d that in some world compatible with our information has the property P. The present definition reads:

$$s[\exists x\phi] = \cup_{d\in D}(s[x/d][\phi])$$

Updating with $\exists x\Diamond Px$ outputs as possible values of x only those d such that in some w compatible with the information in s, d has the property P in w. If $\Diamond Px$ is within the scope of $\exists x$, the consistency test is performed one by one for each $d \in D$, and those d are eliminated as possible values for x for which the test fails.[4]

17.4 Identity and Identification

Consider the following example:

(18) Someone has done it. It might be Alfred. It might not be Alfred.

$\exists! x Px \wedge \Diamond(x = a) \wedge \Diamond(x \neq a)$

(19) It is not Alfred. It is Bill.

$(x \neq a) \wedge (x = b)$

The sequence of sentences (14) is coherent, and hence consistent. If it is continued with (15), everything remains consistent. But viewed as a single utterance, (14) followed by (15) would be incoherent.

There are several situations in which (14) can be coherently asserted. One is the situation in which the speaker is acquainted with the person who

[4]It is these two features which distinguish the present system from the one defined in van Eijck and Cepparello to appear.

did it, but does not know his name — his name might be Alfred, his name might not be Alfred. However, also the opposite case, in which the speaker does know perfectly well who is called Alfred, is possible. In that case the sentence reports that the question is still open whether or not this person did it. A typical example of a situation like this, not involving a name but a deictic pronoun, is this:

(20) Someone has done it. It might be you. But it might also not be you.

$$\exists! x Px \wedge \Diamond(x = you) \wedge \Diamond(x \neq you)$$

Which is consistent and coherent, as you probably would like it to be.

17.4.1 Identification and Identifiers

The following two sentences are consistent:

(21) $\exists! x Px \wedge \forall y \Diamond(x = y) \wedge \forall y \Diamond(x \neq y)$

(22) $\forall x \Diamond(x = a) \wedge \forall x \Diamond(x \neq a)$

These sentences express ultimate forms of non-identification If an information state supports (17), it is known that just one object has the property P, but not which object it is. If an information state supports (18), it is not known of which object a is the name.

Sometimes more information is available.

Definition 16 Let α be a term, $s \in S$.

1. α is an *identifier in s* iff $\forall i, i' \in s: i(\alpha) = i'(\alpha)$
2. α is an *identifier* iff $\forall s: \alpha$ is an identifier in s.

If a term α is an identifier in s, then s contains the information who α is (in at least some sense of *knowing who*). If α is not an identifier in s, then there is at least some doubt about who α is.

Whether or not a term is an identifier in an information state can be tested:

Observation 11

1. *α is an identifier in s iff s supports* $\exists x \Diamond x = \alpha \wedge \forall y(\Diamond y = \alpha \rightarrow y = x)$
2. *α is an identifier iff* $\models \exists x \Diamond x = \alpha \wedge \forall y(\Diamond y = \alpha \rightarrow y = x)$

Identifiers are epistemically rigid designators:

Observation 12 *Let α and β be identifiers.*

1. $\models \Diamond(\alpha = \beta) \rightarrow (\alpha = \beta)$
2. $\models (\alpha = \beta) \rightarrow \Box(\alpha = \beta)$

17.4.2 Why Identifiers are Needed

Identifiers are needed. Otherwise, if we are ignorant at the start, we can never really find out who is who, in the sense of coming to know the names of the objects we are talking about. The following definition and observation explain why this is so.

Definition 17 Let $\langle r, g, w\rangle \in I, \langle r, g', w'\rangle \in I$.
$\langle r, g, w\rangle \simeq \langle r, g', w'\rangle$ iff there exists a bijection f from D onto D such that:

1. For every peg m in the domain of g: $g'(m) = f(g(m))$
2. For every individual constant a: $w'(a) = f(w(a))$
3. For every n-place predicate P:
   $\langle d_1, \ldots, d_n\rangle \in w(P)$ iff $\langle f(d_1), \ldots, f(d_n)\rangle \in w'(P)$

Observation 13 *Let* $\mathbf{0}$ *be the minimal information state.*
If $i \in \mathbf{0}[\phi_1]\ldots[\phi_n]$, *then for every* $i' \simeq i$, $i' \in \mathbf{0}[\phi_1]\ldots[\phi_n]$.

What this observation says is this. If we start out from a state of ignorance — in which names are not identifiers — then, no matter how much information is communicated to us by purely verbal means, we will never get to know to which particular object a given name refers, or which particular objects have which properties. To get this kind of information about the world, purely linguistic means are not sufficient. For identification we need in addition non-linguistic sources of information, such as observation.

To satisfy this need, deictic demonstratives are added to the inventory of the language. It is assumed (rather naively) that if a demonstrative is used, an object is observably present in the discourse situation, which can unambiguously be pointed out to the hearer by the speaker.

Definition 18

1. Let $d \in D$. Then $this_d$ is a term
2. Let $i \in I$. Then $i(this_d) = d$

By definition, demonstratives are identifiers. Once they are added to the language, observation 13 made above, no longer holds. Expressions such as $this_d = a$ are now available, which can tell us which object a refers to.

Identifiers have a special logical role. Suppose the domain consists of two distinct individuals d and d'. We update the state of total ignorance with the following sentence.

(23) $(a \neq b)$

The resulting information state, s, supports

(24) $\Diamond(this_d = a) \wedge \Diamond(this_d = b)$

But s does not support:

(25) $\forall x \Diamond(x = this_d)$

Actually, s does not even allow (21), despite the fact that s supports the two instantiations with a and b — even though these are the names of all the objects around!

The state s does support (22) and (23):

(26) $\Diamond(this_d = a) \wedge \Diamond(this_{d'} = a)$

(27) $\forall x \Diamond(x = a)$

However, at the same time the state s is inconsistent with:

(28) $\Diamond(b = a)$

which can be straightforwardly derived from (23) by universal instantiation — or so it would seem. In other words, universal instantiation is not always valid. In particular, things may go wrong if one instantiates with a term which is not known to be an identifier. Likewise, existential generalization sometimes fails:

(29) $\forall y \Diamond(y \neq a) \not\models \exists x \forall y \Diamond(y \neq x)$

Here, too, generalization is not allowed because the constant a is not an identifier.

17.5 Prospects

We end by indicating some extensions of the system, restricting ourselves to issues which are immediately relevant to the topic of modality and coreference.

In the present paper, the assumption is made that the language users know which objects constitute the domain of discourse. Lifting this fixed domain assumption is relatively easy. The only important change that has to be made, is to keep track of the different stages of the ongoing interpretation process in an explicit manner. This is needed to enable accommodation of existential presuppositions of names and demonstratives in the proper way.[5] In most cases, presuppositions are not accommodated locally, but globally.

For the same reason, this additional structure is needed if the language is extended with anaphoric and non-anaphoric definite descriptions. Accessibility of the various stages of the interpretation process is also indispensable for an account of modal subordination.[6] Subordinated states should remain accessible, since subsequent sentences in the discourse may relate to those, rather than to the global, superordinate level. An analysis of this and other forms of subordination makes it possible to account for the fact that, under certain circumstances, quantifiers inside the scope of negation and the *might*-operator can bind occurrences of variables outside.

[5] For example, along the lines of Zeevat 1992.

[6] See Roberts 1989.

In the present paper, only epistemic modalities haven been investigated. Although the information based nature of dynamic semantics may suggest otherwise, this is not a principled limitation. Alethic modalities can be added, making it possible to implement the Kripkean distinction between metaphysical and epistemic necessity. For this purpose a set of metaphysically possible worlds is added to each possibility. Different possibilities may contain different alternative sets of such worlds. In this way, one can account for the learnability of what is metaphysically possible, necessary, and impossible.

References

Dekker, Paul. 1993. *Transsentential Meditations. Ups and Downs in Dynamic Semantics.* Doctoral dissertation, ILLC/Department of Philosophy, University of Amsterdam, Amsterdam.

Groenendijk, Jeroen, and Martin Stokhof. 1990. Two Theories of Dynamic Semantics. In *Logics in AI*, ed. Jan van Eijck. 55–64. Berlin: Springer.

Groenendijk, Jeroen, and Martin Stokhof. 1991. Dynamic Predicate Logic. *Linguistics and Philosophy* 14(1):39–100.

Roberts, Craige. 1989. Modal Subordination and Pronominal Anaphora in Discourse. *Linguistics and Philosophy* 12(4):683–722.

van Eijck, Jan, and Giovanna Cepparello. to appear. Dynamic Modal Predicate Logic. In *Dynamics, Polarity and Quantification*, ed. Makoto Kanazawa and Chris Piñon. Stanford: CSLI.

Veltman, Frank. 1990. Defaults in Update Semantics. In *Conditionals, Defaults, and Belief Revision*, ed. Hans Kamp. Edinburgh: CCS. Dyana deliverable R2.5.A.

Vermeulen, Kees. 1994. *Explorations of the Dynamic Environment.* Doctoral dissertation, OTS/Utrecht University, Utrecht.

Vermeulen, Kees. to appear. Merging Without Mystery. Variables in Dynamic Semantics. *Journal of Philosophical Logic.*

Zeevat, Henk. 1992. Presupposition and Accommodation in Update Semantics. *Journal of Semantics* 9(4):379–412.

18

Euler and the Role of Visualization in Logic

ERIC HAMMER AND SUN-JOO SHIN

Introduction

In the evolution of diagrams in logic, the participant causing the most spectacular and varied discussion has been Euler diagrams. Euler developed his diagrammatic system in the 18th century for the purpose of providing a visually powerful method of representing syllogisms. Venn was the first to modify Euler's system, in the process developing his own version of the diagrams. Peirce, almost one hundred years later, suggested a variation on Venn diagrams. This modified version of Venn diagrams has been adopted by many elementary logic books and several computer programs. Recently, further research has been carried out on Venn diagrams: to represent generalized quantifiers with them, to formulate and prove the soundness and the completeness of their logic (Shin 1995), to combine them with a first-order language (Hammer 1994), to extend this system to be equivalent to a monadic language in expressive power, and so on.

We claim that the subsequent developments brought about by Venn and Peirce have been motivated by one specific aim, *the aim of increasing the expressive power of Euler diagrams.* This aim, we agree, is important, having been the main motivating force underlying the history of Euler and Venn diagrams. However, we argue that it cost Venn and Peirce certain important features of Euler diagrams. Thus, in our paper we do the following:

(1) We present the history of Euler diagrams from Euler's original version through the modified systems developed by Venn and Peirce.

(2) We point out a negative side to these modifications: while they increase

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

the diagrams' expressive power, they lose a good deal of the visual clarity that make Euler diagrams so valuable in the first place. This lesson also has ramifications for the role of diagrams in logic generally. In particular, it emphasizes the importance of retaining the visual power of one's diagrams even if it means giving up expressive power, flexibility, generality, or formal advantages.

(3) We analyze exactly how and why Euler diagrams are more visually powerful than the variations introduced by Venn and Peirce.

(4) We reconstruct Euler's system to overcome certain legitimate problems pointed out by Venn and Peirce without sacrificing visual clarity. This reconstructed version of Euler diagrams is given a precise semantic analysis. Rules of inference for manipulating the diagrams are also given.

(5) Soundness, completeness, and decidability results are proved for the system.

(6) We discuss some problems concerning the relationship between restrictions put on the syntax of the diagrams and the expressive power of the resulting system.

In the abstract we briefly outline these matters.

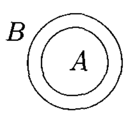
18.1 Euler Diagrams

Euler uses a drawn circle tagged with a letter, say 'A', to represent a set A. The circle divides the page into two regions, the region inside the circle representing the set A and the region outside the circle representing the complement of A. Euler's use of drawn circles to represent sets is governed by the following basic convention:

Basic Convention *Every object x in the domain is assigned a unique location, say l_x, in the plane such that l_x is in region R if and only if x is a member of the set region R represents.*

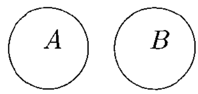
So each member of the set A is assigned a location within the circle labeled 'A' while each object that is not a member of A is assigned a location outside of the circle.

A useful consequence of Euler's idea is that the meanings of more complicated diagrams, diagrams involving more than one circle, fall out of his basic convention. Consider the following diagram asserting that all A's are B's:



Take any object x in A. By the basic convention x is assigned a location within the circle labeled 'A'. Therefore, by the physical properties of the insideness relation, x is assigned a location within the circle labeled 'B'.

So by the basic convention x is also a member of B. A similar argument shows that the following diagram asserts that no A is B:

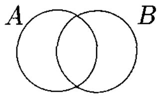


In Euler's system, then, the domain is represented by locations on the page while sets are represented as drawn circles on the page separating locations within the circle from locations outside of the circle, so that relationships among sets are represented by relationships among the drawn circles. Some of the power of the system lies in the fact that an object being a member of a set is easily conceptualized as the object falling inside the set, just as locations on the page are thought of as falling inside or outside drawn circles. The system's power also lies in the fact that no additional conventions are needed to establish the meanings of diagrams involving more than one circle: relationships holding among sets are asserted by means of the same relationships holding among the circles representing them.

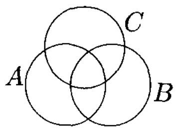
There are two real problems with Euler's system, aptly criticized by Venn and Peirce, which we can only mention in the abstract. The first concerns Euler's mechanism for making existential statements such as 'some A is not B' by means of the particular placement of the letters tagging the circles. The conventions governing this mechanism suffer from serious problems of ambiguity. The second problem with the system is that one cannot express partial information about the relationships among the sets represented in a diagram. For example, one cannot assert that $A \subseteq (B \cup C)$ with a diagram without also committing oneself to some particular relationship holding between B and C: that all B are C, that all C are B, that no B is C, that some B is C, that some B is not C, or whatever.

18.2 Venn's and Peirce's Revisions

In seeking to overcome Euler's inability to express partial knowledge about the relationships between sets, Venn gave up Euler's fundamental idea of representing relationships among sets by using corresponding relationships among the drawn circles representing them. Rather, Venn required circles to be drawn so that they overlapped with the other circles in every possible combination. He then had to introduce an extra syntactic device, shading, to assert that the emptiness of a represented set. Thus, for Venn, *every* diagrammatic assertion about two sets A and B is based on the diagram



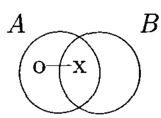
while every assertion about three sets A, B, and C is based on the diagram:



Similarly for diagrams involving four or more closed circles. The two diagrams above say nothing about the relationships between the sets represented. To make a claim one must add shading to the appropriate region. For example, the shading of the overlap of the A-circle and B-circle in either diagram would be an assertion that $A \cap B$ is empty.

Venn retains the ability to represent partial information about the sets represented, then, in the following way. He represents all possible combinations of the circles, yet by merely representing a combination he means to assert nothing about whether the set it represents is empty or not. The only way to make an assertion is to shade a region, thereby asserting that the set it represents is empty.

Peirce adopts the same strategy as Venn, the only difference being that he uses a different syntactic device than shading to make assertions about the various sets represented. In particular, he uses the symbol 'o' to assert the emptiness of a set (just like Venn's shading). He also uses the symbol 'x' to assert non-emptiness. Finally, he allows one to disjoin x's and o's using a bar '——'. For example, the diagram

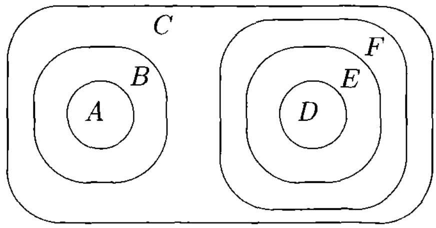


makes the claim: all A are B, or else some A is B.

While Peirce's and Venn's modified systems allow for the representation of partial information about the relationships between sets, they suffer from some serious drawbacks.

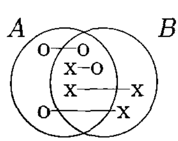
First, both Venn and Peirce need to introduce additional syntactic devices to make assertions about the relationships between sets, while Euler does not. By relying on his basic convention, he is able to use relationships among circles to make claims about the relationships among the sets they represent.

Second, there is a serious loss of visual clarity in the transition from Euler's original diagrams to Venn and Peirce diagrams. This becomes particularly clear when the number of circles in a diagram grows beyond three. Syntax quickly becomes a jungle with the Venn and Peirce diagrams, but for certain kinds of information remains quite perspicuous with the Euler diagrams. For example, the following Euler diagram represents relationships between six sets:



The corresponding Venn or Peirce diagram asserting the same thing is practically impossible to draw, let alone comprehend. With a Venn or Peirce diagram having n curves, there are $2^n - 1$ distinct compartments. For example, with 6 curves there are 63 enclosed regions in a Venn or Peirce diagram, compared to the mere six enclosed regions in the Euler diagram above.

The visual power of Euler diagrams is also a product of his use of relationships among circles to represent relationships among sets, rather than relying on new syntactic devices like shading or the symbol 'o'. It is quite easy to think about containment and disjointness relationships among sets in terms of containment and disjointness relations among the circles representing them. The same is not true with the new syntactic devices of Venn and Peirce. For example, Peirce gives the following diagram as an example:



The meaning of this diagram is certainly not perspicuous the way a good diagram should be. As it turns out, it asserts: either all A are B and some A is B, or else no A is B and some B is not A. With Venn's and Peirce's revisions, most of Euler's original ideas about the power and clarity of the visual representation of assertions about sets were forgotten except that a geometrical object, the circle, is used to represent sets.

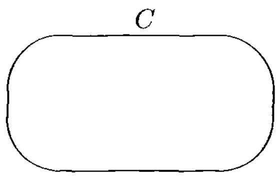
18.3 A Reconstruction and Semantic Analysis of Euler's System

In our reconstruction of Euler's original system we do three things: (1) we drop Euler's mechanism for making existential claims; (2) we retain the ability to represent partial information about the relationships between sets, as Venn and Peirce desired; and (3) we retain Euler's original and crucial idea of using relationships among circles to represent the corresponding relationships among the sets they represent.

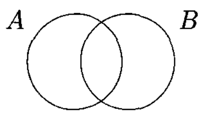
For the syntax of the diagrams one needs only circles, which can be drawn together in any arrangement, and labels, so that each circle of a given diagram can be tagged by a unique label.

The basic idea behind the semantics of the diagrams is that if some set is intentionally not represented in a diagram (such as $A \cap B$ or $A - B$ or $(A \cup B) - C$), then one has thereby asserted that it is empty. However, if some set *is* represented, such as the set represented by the overlap of two circles, we do not take that to assert anything in particular: the intersection may or may not be empty; nothing is being said one way or another. Thus, *by dividing up the page with some arrangement of a diagram's circles, we assert that the domain can be divided up in the same way among the sets represented by the circles.* Like Venn, simply by representing a set we are saying nothing one way or the other about it. Unlike Venn, however, by violating the requirement that each circle overlap the others in all possible combinations, we can make claims about how the domain can be divided up among the sets the circles represent.

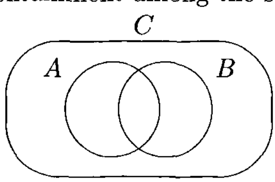
As an example, suppose one wanted to assert that $(A \cup B) \subseteq C$ by means of an Euler diagram. To represent C one would one would use a single circle:



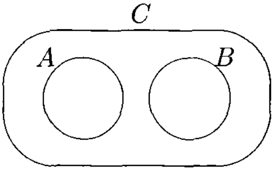
To represent $A \cup B$ one would draw two circles labeled appropriately, and since one is asserting nothing about the relationship between A and B, one would draw them so to represent all possible combinations of the two sets:



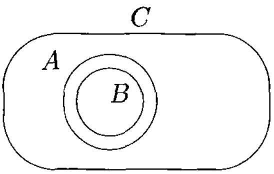
To assert that $A \cup B$ is a subset of C, we simply draw the diagram representing $A \cup B$ within the circle representing C, using containment among circles to represent containment among the sets they represent:



On the other hand, the diagram



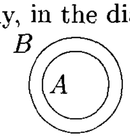
also asserts that $A \cap B = \emptyset$ since there is no region representing $A \cap B$, that is, there is no area of overlap of the circle labeled 'A' and the circle labeled 'B'. So it is equivalent to the sentence $((A \cup B) \subseteq C) \wedge (A \cap B = \emptyset)$. Likewise, the diagram



also asserts that $B \subseteq A$, so is equivalent to the sentence $((A \cup B) \subseteq C) \wedge (A \cap B = \emptyset)$, that is, to: $B \subseteq A \subseteq C$.

18.4 Semantics

For the semantics we need the concept of a 'minimal region' of a diagram. Consider the previous diagram. It has four minimal regions: the region within all three circles, the (donut-shaped) region within the two circles labeled 'A' and 'C' but outside the circle labeled 'B', the region within the circle labeled 'C' but outside the other two circles, and the region outside all three circles. Similarly, in the diagram



there are three minimal regions: the one within both circles, the one within neither circle, and the one within the circle labeled 'B' but outside the circle labeled 'A'. In general, a region r of a diagram D is a 'minimal region' if and only if there are sets of D's circles In and Out such that $In \cap Out$ is empty, $In \cup Out$ is the set of all of D's circles, and r is the region within each circle in In but outside of each circle in Out. (Either In or Out could be empty.)

Let U be a set. Then a function f is called a 'U-function' if and only if it assigns a subset of U to each circle, with the only restriction being that it assigns the same subset to any two circles sharing the same label. Intuitively, f is a model of our intention to represent certain subsets of the domain with the various circles of a diagram.

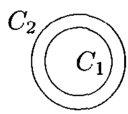
With the concepts of minimal region and U-function we can now define model and truth in a model. A model will be understood as assigning subsets to minimal regions as follows:

Definition 1 A pair (U, I) is a 'model' if and only if (1) U is a set, (2) I is a function assigning subsets of U to minimal regions, and (3) there is a U-function f such that if m is a minimal region of a diagram D, $c_1, \ldots, c_m$ are the circles of D m falls within and $c_{m+1}, \ldots, c_n$ are the circles of D m falls outside of, then $I(m) = (f(c_1) \cap \ldots \cap f(c_m)) - (f(c_{m+1}) \cup \ldots \cup f(c_n))$.

Whether a diagram is true in a model depends, in essence, on whether the domain can be divided up among the sets represented by the diagram's circles the way the circles divide up the locations on the page. In particular, it is true in a model if and only if every object falls within the set assigned to one of the diagram's minimal regions:

Definition 2 Let D be a diagram and (U, I) be a model. Then 'D is true in (U, I)' if and only if for all $x \in U$ there is a minimal region m of D such that $x \in I(m)$.

To illustrate the definition, consider the following diagram:



Let U be a set and f be a U-function assigning some set C_1 to the circle labeled 'C_1' and some set C_2 to the other one. Then the model based on U and f will assign $C_1 \cap C_2$ to the minimal region within both circles, $C_2 - C_1$ to the donut-shaped minimal region, and $U - (C_1 \cup C_2)$ to the minimal region outside of both circles. The diagram is true in the model if and only if each object in U falls into one of these three subsets of U. This holds if

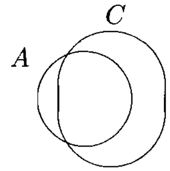
Definition 3 Let $\Delta \cup \{D\}$ be a set of diagrams. Then 'D is a logical consequence of Δ' if and only if for all models (U, I) in which every member of Δ is true, D is true.

The proof of the following theorem shows that the logical consequence relation can be characterized by a mechanical procedure:

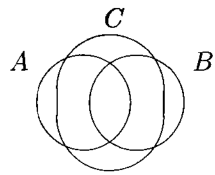
Theorem 1 (Decidibility) *There is a decision procedure for determining of any finite set $\Delta \cup \{D\}$ of diagrams whether or not D is a logical consequence of Δ.*

18.5 Rules of Inference

The rule 'Erasure' is a simple rule that allows one to erase any circle from a diagram. For example, it allows one to infer



from

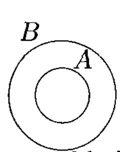


The rule is valid, intuitively, because if the domain can be divided up among sets the way a diagram divides up the page among the circles representing the sets, then it can be divided up the way the result of erasing one of the circles divides up a page, since one is then simply eliminating the need to consider one of the sets.

Another rule, 'Introduction of a New Circle' allows one to add a circle to a diagram in such a way that nothing additional is thereby asserted (i.e., so that it overlaps a proper part of each minimal region of the diagram):

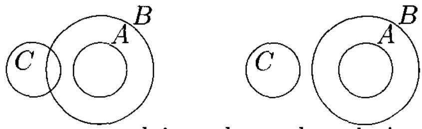
Introduction of a New Circle One can add a circle tagged by a new label to a diagram provided that it overlaps a proper part of each minimal region of the diagram.[1]

For example, suppose one wanted to add a circle labeled 'C' to the following diagram:

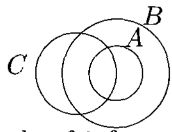


It would be wrong to add the circle to overlap the B-circle but not the A-circle, because the new diagram would then assert that A and C were disjoint, something that we do not know to be the case. Likewise, it would be wrong to add the new circle so to overlap neither the A-circle nor the B-circle, for that would be to assert that B and C were disjoint. So the following two options are incorrect:

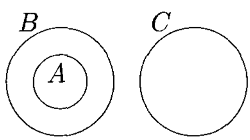
[1]For this rule to be applicable to every diagram, one must liberalize the notion of 'circle' to include non-convex closed curves.



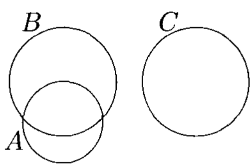
To avoid making any new claims about the relationships between the new set and the original ones as with the above, the new circle must be drawn so as to overlap each minimal region of the diagram, but only a proper part of each minimal region. Thus, the new circle should be added as in the following:



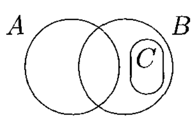
The final rule of inference for reasoning with diagrams allows one to change the arrangement of the circles of a diagram provided that one weakens rather than strengthens the assertion one is making. Since a diagram makes an assertion about how the domain can be divided up by the way its circles are arranged, it is valid to rearrange the placement of one of the circles so to allow for more possibilities. For example, from the diagram



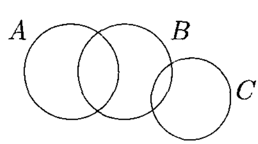
it is legitimate to move the A-circle to get:



This is a valid inference because one has done nothing but weaken one's assertion about how the domain can be carved up. Similarly, from the diagram



it is legitimate to redraw the circle labeled 'C' as follows:



The original diagram says that C is a subset of $B - A$, while the second says that C and A are disjoint. One has simply redrawn the one circle so to allow for additional possibilities in dividing up the domain; in particular, one has added the possibility that some objects are in C but neither B nor A.

To state the rule "Weakening' rigorously, we need the concept of two minimal regions being counterparts. Suppose m and m' are minimal regions of two diagrams D and D', respectively. Then m and m' are 'counterparts' if and only if there are circles $C_1, \ldots, C_m$ and $C_{m+1}, \ldots, C_n$ of D and circles $C'_1, \ldots, C'_m$ and $C'_{m+1}, \ldots, C'_n$ of D' such that (1) for each i, $1 \leq i \leq n$, C_i and C'_i are tagged by the same label; (2) m is the minimal region within $C_1, \ldots, C_m$ but outside of $C_{m+1}, \ldots, C_n$; and (3) m is the minimal region within $C'_1, \ldots, C'_m$ but outside of $C'_{m+1}, \ldots, C'_n$.

Given this definition of two minimal regions being counterparts, the following is a statement of the rule:

Weakening D' results from D by this rule if and only if (1) D and D' have the same number of circles tagged by the same letters, and (2) every minimal region of D has a counterpart in D' (though not necessarily vice versa).

Proofs involving only diagrams can now be carried out by simply stringing together applications of the above rules. In other words, a diagram D' is 'provable' from a diagram D (written $D \vdash D'$) if and only if there is a sequence of diagrams $D, D_1, \ldots, D_n, D'$ such that each member of the sequence follows from previous members of the sequence by one of the three rules of inference: Erasure, Introduction of a New Circle, and Weakening.

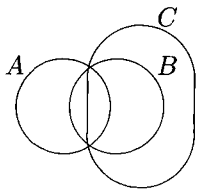
We have the following soundness and completeness theorem linking provability with logical consequence:

Theorem 2 (Soundness/Completeness) *Let D and D' be Euler diagrams. Then $D \vdash D'$ if and only if D' is a logical consequence of D.*

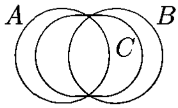
18.5.1 Syntactic Stipulations and Expressive Power

Euler diagrams have an interesting feature that appears to have no counterpart in standard symbolic systems, a feature concerning the way syntactic constraints limit expressive power. In particular, there are a variety of stipulations one can put on the syntax of Euler diagrams, each of which results in a system with a different expressive range. Unlike with symbolic systems, however, there seems to be no principled way to choose between these various systems.

Some examples will make the matter more clear. Suppose one stipulates that no three circles can intersect at the same point. The following diagram will therefore be non-well-formed

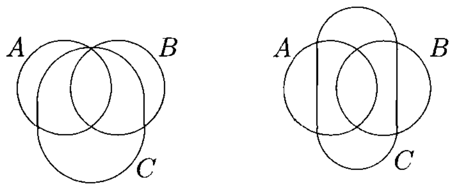


and hence one will be unable to assert that $(B - A) \subseteq C$. Likewise, the diagram

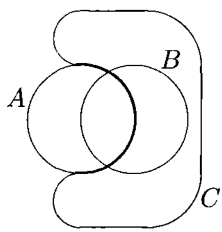


will also be ill-formed, so that one will be unable to express that $C \subseteq (A \cup B) \wedge (A \cap B) \subseteq C$.

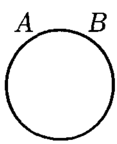
If one stipulates that well-formed diagrams must be such that no three circles can intersect at the same point and that no two distinct minimal regions of the same diagram can be counterparts, then one will be unable to express that $(A \cap B) \subseteq C$ because both of the following diagrams will be ill-formed:



Likewise, if one allows two circles to intersect at infinitely many points one will be able to assert that $(B - A) \subseteq C \wedge A \cap C = \emptyset$ by means of the following diagram:



One would also be able to assert that $A = B$ by drawing the two circles one on top of the other as follows:



However, each of these claims would be inexpressible if one were to not allow two circles to intersect at infinitely many points.

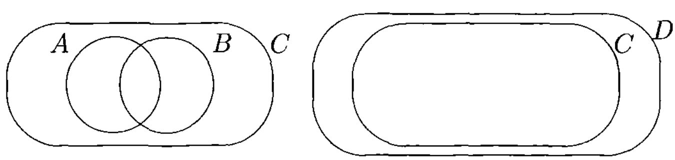
One also gets variation in expressive power depending on the stipulations about whether or not the "circles" must be convex, whether they must be differentiable, whether or not they may self-intersect, and so forth. There is no obvious best decision in defining the syntax here. The only motive is not to get as much expressive range as possible, but to get as much as is possible yet still have clear, simple, readable diagrams.

Our rules of inference and decision procedures are mostly independent of these sorts of topological decisions concerning the syntax of well-formed diagrams. Whatever stipulations are adopted, the notions of minimal region and counterpart will remain the same. The only real connection between these stipulations and our rules is that we need the rule of Introduction of a New Circle to work out; that is, we need our diagrams to have the property that for any well-formed diagram, one can add a new circle that overlaps a proper part of each minimal region of the diagram.

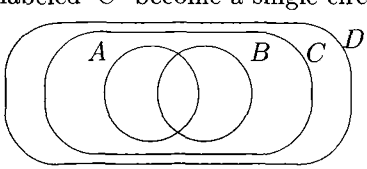
Given that expressive power varies with the syntactic stipulations one adopts, a problem is to describe a set of monadic first-order sentences having the same expressive power as the set of Euler diagrams given certain stipulations. While this is quite easy in the case of Venn diagrams, it is not so obvious in the case of Euler diagrams.

18.5.2 Unification

Sometimes one would like to combine together or unify two diagrams into a single diagram that carries the information of both, their 'conjunction'. One might, for example, want to unify together the following two diagrams:



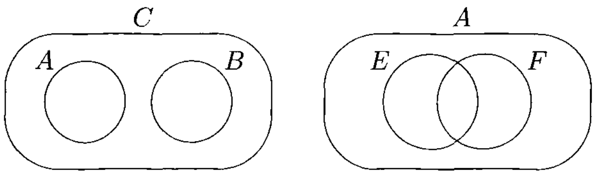
Their common element is having a circle labeled 'C'. The first one has every other circle falling within its circle labeled 'C', while the second one falls has circles containing its circle labeled 'C' but none falling within it. To unify the two diagrams one must take advantage of this common element. Thus, one simply overlays the one diagram over the other in such a way that the two circles labeled 'C' become a single circle:



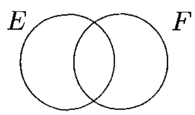
One has thus carried out an almost physical unification, the sort of process that could be carried out with scissors, literally cutting out the material within the circle labeled 'C' in the first diagram and pasting it within the circle labeled 'C' in the second diagram (possibly reducing or enlarging as needed).

A restricted form of unification as in the above example, then, is as follows. *Let D and D' be diagrams such that for some label x, D has a circle labeled by x such that every other circle of D falls within it, and D' also has a circle labeled by x which no other circle falls within. Then one can infer a new diagram, the 'unification' of the two, which is the result of physically cutting out whatever falls within D's circle labeled x and pasting it into D''s circle labeled x, reducing or enlarging as necessary.*

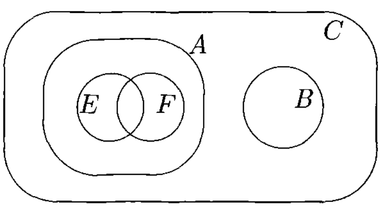
For another example of this rule, suppose one wanted to unify the following two diagrams:



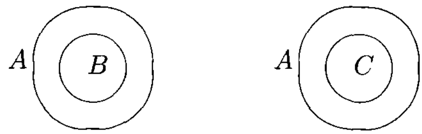
Each of them has a circle labeled 'A'. In the first one, nothing falls within the circle, while in the other every other circle falls within it. Therefore, one simply transfers whatever falls within the circle labeled 'A' in the second diagram, namely



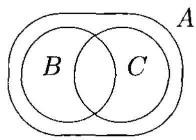
to the area within the circle labeled 'A' in the first diagram to get:



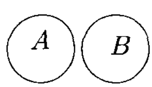
The above version of unification seems to be very visually appealing. One can conceptualize it as a very physical process, as consisting of removing the material from one bounded area and transferring it to another bounded area. However, it is also a special case of the more general operation of unifying together two arbitrary diagrams. It requires that one find a label common to the two diagrams, one of which tags a circle having nothing within it, the other one tagging a circle having every other circle within it. A more general form of unification would be needed to handle cases such as unifying together the following two diagrams:



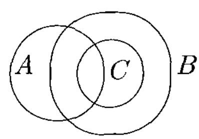
Intuitively, in the unification a single circle labeled 'A' should contain a circle labeled 'B' and a circle labeled 'C'. Moreover, since we know nothing about the relationship between B and C, these should be drawn so to partially overlap:



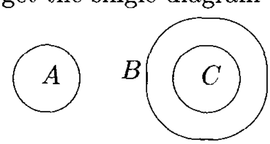
Likewise, the two diagrams



and

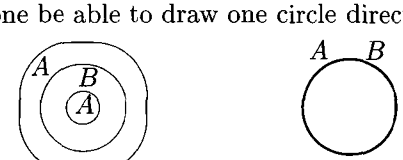


should be unifiable to get the single diagram as follows:



However, there is a serious obstacle to providing a general unification rule, a rule that allows one to infer from any two diagrams D_1 and D_2 a diagram D whose models are exactly those that model both D_1 and D_2. The problem is that unlike our three rules of inference given above, unification is much more closely tied up with the particular syntactic stipulations one adopts. The simplest example of this is seen with the following two diagrams:

There are two ways one might attempt to unify them. One requires the same letter to tag two different circles in the same diagram while the other requires that one be able to draw one circle directly over another one:



Problems such as these it uncertain that a general unification rule should be a priority. If it requires us to define the syntax in such a way that diagrams lose the clarity and simplicity that make them preferable to later revisions in some circumstances, then it seems that the rule should be abandoned in favor of the more fundamental consideration of visual power. There is an interesting question, though, as to whether a reasonable set of syntactic stipulations can be given for which a general unification rule is possible, a rule that, along with the other three rules above, is strongly complete.

18.6 Conclusion

Venn and Peirce noticed that Euler's original system had two main flaws: an ambiguous device was used to make existential claims, and the system did not allow one to remain uncommitted about relationships between the sets represented. Venn and Peirce managed to improve this situation, but in the process they also lost a key element of Euler's system, the element that made it so visually appealing. In particular, they gave up Euler's method of representing relations among sets by means of drawing analogous relations among drawn circles. In our revised version of Euler's original system we attempted to avoid the problematic aspects of the original system while retaining its power and spirit. While this necessitated foregoing the gain in expressive power found in Venn's and Peirce's systems, the compensation was an increase in visual clarity.

References

Euler, Leonhard. 1986. *Briefe an eine deutsche Prinzessin.* Friedrich Vieweg and Sohn.

Hammer, Eric. 1994. Reasoning with Sentences and Diagrams. *The Notre Dame Journal of Formal Logic* 34.

Shin, Sun-Joo. 1991. A Situation-Theoretic Account of Valid Reasoning with Venn Diagrams. In *Situation Theory and Its Applications, Vol. 2.* Stanford: CSLI.

Shin, Sun-Joo. 1994. Peirce and the Logical Status of Diagrams. *History and Philosophy of Logic* 15.

Shin, Sun-Joo. 1995. *The Logical Status of Diagrams.* Cambridge University Press.

Venn, John. 1971. *Symbolic Logic.* Burt Franklin. 2nd edition.

19

Individuation and Reasoning

KIYOSHI ISHIKAWA

Introduction

Since Donnellan (1966), the referential/attributive (R/A) distinction has attracted wide attention. According to Donnellan, a definite description is used referentially when the speaker has a particular individual in mind (otherwise it is used attributively). Let's call this the particular individual conception of the R/A distinction. On the other hand, he also assumes that the descriptive content of a given definite description has different status depending on whether it is used referentially or attributively. Let's call this the descriptive content conception. Many people have assumed that these two conceptions give the same result. However, as Recanati (1981) and Kronfeld (1986) have already noticed and as we will see below, this is wrong, at least if we are to understand the latter in an intuitive way.

Contrary to Recanati, we construct an information-based theory of the R/A distinction based on the descriptive content conception and give an analysis of the particular individual conception in terms of it. In the picture of semantics that we get, "meaning" is analyzed in terms of information-state changing potential for cognitive agents,[1] rather than truth condition with respect to an agent-independent model.[2]

This research was conducted while I was a Fulbright scholar at the University of California, Santa Cruz I would like to thank Robin Cooper, Masahito Kawamori, Yasuhiro Katagiri, Kuniaki Mukai, Bill Ladusaw, John Perry, Stanley Peters, Atsushi Shimojima, Hidetosi Sirai, Hiroyuki Suzuki, Syun Tutiya for discussion and/or comments Of course, none of them necessarily agrees with my ideas

[1]Cf Heim's (1982) file change potential

[2]The proposed theory can be seen as a technical refinement of Shimojima's (1993) theory (see also Fauconnier 1985) However, whether Shimojima himself regards this as a refinement or not is another issue Roughly speaking, our theories have some

Logic, Language and Computation
Jerry Seligmann and Dag Westerståhl, eds

19.1 The Scenarios

19.1.1 The First Scenario

Consider the following simplified version of Shimojima's (1993) scenario.

SCENARIO 1
Here we are interested in what Watson believes based on Holmes's utterance. He believes whatever Holmes says.

STAGE 1
In front of Smith's mutilated body, they are wondering who murdered him. Observing Smith's body, Holmes utters (1).

(1) Smith's murderer is insane.

Later Holmes says (2).

(2) Jones is Smith's murderer.

Further, upon interviewing Jones, he reports his finding by uttering (3).

(3) Smith's murderer is a Buddhist.

Now Watson believes that Jones is insane and a Buddhist.

STAGE 2
Later Holmes says (4).

(4) Bond is Smith's murderer (not Jones).

Still Watson believes that Jones is a Buddhist, but he doesn't have to believe that Jones is insane. Instead, he will believe that Bond is insane.

OBSERVATION
Thus, murdering and insanity necessarily have to be predicated on the same single individual. These two properties "go together." However, murdering and Buddhism do not necessarily have to "go together."

(1) and (3) are paradigm cases of A-use and R-use, respectively, and it seems intuitively clear that the above patterns of reasoning are due to the

affinity with Kripke (1979) and Searle (1979). See Ishikawa (1995) for a comparison between these theories.

R/A distinction In this case, the two conceptions of the R/A distinction give the same result Thus, this scenario could be explained simply by regarding the R/A distinction as a singular/general distinction, in which case the utterance (1) would convey something like (5a) while the utterance (3) would convey something like (5b) [3]

(5) a `(the` x `murderer-of-Smith(`x`))(insane(`x`))`
  b `Buddhist(Jones)`

From (5a) it follows that, if Jones is Smith's murderer, then Jones is insane, but if Bond is Smith's murderer, Bond is insane However, (5b) only says that Jones is a Buddhist, whether he is Smith's murderer or not Thus, under the assumption that the R/A distinction is a singular/general distinction, the expressed propositions (5a-b) would capture Watson's understanding of Holmes's utterance (1) and (3), and there would be no mystery in Watson's reasoning patterns in Scenario 1

[3]There is a relatively straightforward way to obtain such a difference in Situation Semantics Assume that we employ Gawron & Peters's (1990) *Absorption Princi ple* Let's not worry about definiteness too much, just assume that the relevant resource situations have a unique murderer Then, we can analyze (1) (3) as (i) (ii) respectively

(i) s_d $T1$
where
$T1 = [\ s\ x \mid s \models \ll \text{insane}, x_{s \models \ll \text{murderer of}\ \ x\ \ \text{Smith} \gg} \gg]$

(ii) s_d $T3$
where
$T3 = [\ s \mid s \models \ll \text{buddhist}, y_{m \models \ll \text{murderer of}\ \ y\ \ \text{Smith} \gg} \gg]$

The resource situation m in (ii) may be a situation characterizing Holmes s belief, or s_d as seen by Holmes and Watson Whichever is the case, y in (ii) is a parameter, which has to have been anchored in the interpretation That is (ii) is only a constraint, and given the circumstance of the utterance, which anchors y the inter pretation is (iii) Jones is a constituent of the proposition, but the property of being Smith's murderer is not

(iii) s_d $T3'$
where
$T3' = [\ s \mid s \models \ll \text{buddhist}, \text{Jones} \gg]$

Note that, in such an analysis the singular/general difference is obtained by the effect of the Absorption Principle, which can be triggered even if the resource situation for a given definite description does not coincide with the described situation (that is, if the description's content parameter stands in the transitive closure of the *depends on* relation to the absorbed described situation parameter) Such an analysis is immune to Soames's (1986) criticisms of Barwise & Perry (1983) (for detailed illustration of this analysis, see Ishikawa 1995)

Whether one accepts such an analysis or not, what we are focusing our attention on here is the assumption that the R/A distinction amounts to a singular/general distinction, not on whether the distinction should be obtained by absorption or by some other (semantic or pragmatic) mechanism

However, the following variants of this scenario show that the two conceptions of the R/A distinction do not give the same results.

19.1.2 The Second Scenario

First we consider the following variant of Scenario 1.

Scenario 2

Stage 1
First Holmes says (6), as in Scenario 1.

(6) Smith's murderer is insane.

Then Holmes says:

(7) I interviewed Smith's murderer (but I'm not going to tell you who it is).

Watson asks him his finding from the interview, and he replies:

(8) Smith's murderer is a Buddhist.

Now Watson will believe that there is a single person who is Smith's murderer, insane, and a Buddhist.

Stage 2
Later Holmes says:

(9) I was wrong about who murdered Smith. Another man did.

Now Watson will believe that there is an insane murderer around, in addition to an interviewed Buddhist.

Observation Watson has never fixed his idea as to which real world individual is Smith's murderer, but the same patterns as Scenario 1 obtain. Further, Watson can become a second hand speaker to a third person, giving rise to the same patterns of reasoning in this third person. Thus, R-use does not depend on "knowing which real world individual."

The analysis of Scenario 1 suggested above presupposed the following two assumptions.

Assumption 1: The R/A distinction amounts to a singular/general distinction between the (alleged) expressed propositions.
Assumption 2: The R/A distinction explains Watson's understanding

of Holmes's utterances (and hence Watson's patterns of reasoning)

When it comes to Scenario 2, there is certain tension between these two assumptions, our intuitive understanding of the notion of *identifying a real world individual,* and the usual notion of *singular proposition* Since (8) supports the same pattern of reasoning of Watson as the intuitively referential utterance (3), Assumption 2 dictates that (8) should be regarded as an instance of R-use Then, due to Assumption 1, it follows that Watson's *understanding* of (8) should be analyzed in terms of a singular proposition about a real world individual, even though Watson has not fixed his idea as to which real world individual it is (in an intuitive sense) Hence, we would have to say that "knowledge by description" (in the Russellian sense) can enable one to understand a singular proposition This also suggests that the particular individual conception of the R/A distinction is in conflict with the descriptive content conception The latter dictates that (8) is an instance of R-use, but some people's intuition might resist to the idea that Watson's *understanding* of (8) is about a particular individual

Indeed, if we give up the idea to interpret the notions of *singular proposition* and *particular* epistemologically, that is, if we dissociate these two notions from the intuitive notion of *identifying a real world individual,* it might be thought that the above consideration will not necessarily mean that Scenario 2 is a counterexample to either one of the above assumptions However, we can extend this scenario so as to show that this is wrong When Watson reports that Smith's murderer is a Buddhist, we can say that *Smith's murderer* was used to talk about whoever was interviewed by Holmes So, if I tell him that Jones is the interviewed guy, then he will believe that Jones is a Buddhist But if I tell him later that Bond is the interviewed guy, then he will believe that Bond is a Buddhist Thus, "being interviewed by Holmes" and "being a Buddhist" go together Then, we do not want to characterize Watson's understanding of (8) in terms of a singular proposition, since it supports Watson's reasoning about different individuals, depending on his epistemic state This is in direct conflict with the above set of assumptions [4]

[4]In the absorption based analysis noted above the strategy was to express this kind of link between properties (i e those of being interviewed by Holmes and of being a Buddhist) in terms of situation type and absorption Then, we would want something like (1) or some refined version of it

(1) $[\ s, x \mid s \models \ll \text{buddhist}, x_{s \models \ll \text{interview Holmes } x \gg} \gg]$

However, it does not make sense at all that a murderer parameter is anchored to something absorbed Then, we have to deal with the combination of being interviewed and being a Buddhist in some way other than absorption But if this is possible, then it's not clear what absorption buys us here

19.1.3 The Third Scenario

Next consider the following scenario.[5]

Scenario 3: Referring to "the P" by a Name

Assume that, at Stage 1 of Scenario 1, shortly after saying that Jones is Smith's murderer, Holmes conducts further inspection of Smith's dead body. Watson asks what he has found out. If he replies by saying (10), it will give the same effect as (11).

(10) Jones used belladonna (to kill Smith).

(11) Whoever is Smith's murderer used belladonna (to kill Smith).

At the end of stage 1, Watson will believe that Jones used belladonna to kill Smith. However, at stage 2, when Holmes says that Bond, but not Jones, is Smith's murderer, Watson will believe that Bond used belladonna to kill Smith. Then, *Jones* was used to refer to "whoever is Smith's murderer"!!!

Of course, if Jones is not a murderer, it will not make a good sense to say that he used belladonna. So it's no mystery here that the property of using belladonna will be "removed" from Jones at stage 2, irrespective of what we assume about the R/A distinction etc. But the point is that this property does not simply go away. At stage 2 Bond will be believed to have used belladonna. Thus, "being Smith's murderer," "being insane" and "using belladonna" go together. So we want (10) to contribute something like (12) to Watson's understanding.

(12) $(\texttt{the}\ x\texttt{:}\ \texttt{murderer-of-Smith}(x))(\texttt{used-belladonna}(x))$

Thus, *Jones* can be *used* as a disguised description. However, we do not want simply to *analyze* it as a disguised description—as far as linguistic semantics in our grammar is concerned, we want to *analyze* the name *Jones* as something like (13) and want to clarify how its disguised description use arises.[6]

(13) $x_{r\models\ll \text{named}, x, '\text{Jones}'\gg}$

The above set of assumptions is in direct conflict with Scenario 3. If names are directly referential (in the sense of Kaplan 1989), then (10)

[5] Atsushi Shimojima had also come up with a similar scenario (p.c. 1994).

[6] (13) means a parameter whose restriction says that it is named Jones in its resource situation r.

would convey something like (14) to Watson, where `Jones` is a real world individual.[7]

(14) `used-belladonna(Jones)`

Then, this is a singular proposition, and hence should be classified as an instance of R-use, by Assumption 1. However, Watson's understanding of (10) is "general" as opposed to "singular" in Scenario 3. This is a contradiction. Then, we would have to say that names are not (always) directly referential, in which case we would want to preserve the "descriptive content" of the name as a constituent of the expressed proposition, which is the property of being named Jones (or the property of being this fixed individual Jones, if you like). However, in Scenario 3, such a property does not go together with the property of using belladonna. A contradiction again.

19.2 The Proposed Analysis

19.2.1 Preliminary Considerations

Thus, at least either one of the above assumptions has to be discarded. The analysis proposed here discards Assumption 1 while preserving Assumption 2. Here I only mention two motivations for this decision.

First, we want some account of the R/A distinction anyway. We also want some account of the above patterns of reasoning anyway. Assumption 2 says that the two are explained by the same mechanism. However, if we discard Assumption 2 and preserve Assumption 1, then we need separate mechanisms for the two, which is not economical.

Second, (as far as our understanding of utterances is concerned) *singular proposition* and *particular individual* are not obvious notions. For example, suppose there are identical twins whose family names are Lucia. Ken has seen both of them, but he thinks that they are the same person. Now Ken utters:

(15) Lucia wants to study computer science.

Now, what proposition did Ken express? I see no obvious answer to this question.

19.2.2 Individuals as Uniformities

Suppose Ken and Tom saw somebody yesterday. Also suppose they saw somebody this morning. Ken thinks that they are the same person, while Tom thinks they are not. Then, for Ken, they form a uniformity across the two experiences, but for Tom, they do not. And in general, we regard such uniformities as individuals. Our theory is based on this observation.

---

[7]Whether the possibly understood adjunct *to kill Smith* shows up in the semantics or not is not crucial.

We assume the familiarity conception of definiteness (Heim 1982). For the moment we also ignore who was murdered. We analyze the sentence *Smith's murderer is insane* as giving the constraints in (16).[8]

(16) a. $s_1 \models \ll \text{murderer}, X1 \gg$
b. $s_2 \models \ll \text{insane}, X2 \gg$
c. $s_1 \models \ll \rightarrow, X1, N \gg$
d. $s_2 \models \ll \rightarrow, X2, N \gg$

Because of the definiteness, the information (16a) has to be familiar to the conversation participants. $X1$ and $X2$ are *roles* in s_1 and s_2, respectively. (16c-d) say that they should be in the arrow relation to a single common thing N, or more intuitively, they constitute a uniformity as an individual.[9] When $X1$ equals $X2$, A-use arises. Otherwise, R-use arises. Then, the R/A distinction is not whether a real world individual appears in the propositional content, but rather whether two different roles are involved.

(16) as a whole can be understood as an instruction to update your informational state. $X1$ is already familiar to you. You have to look for an appropriate role $X2$ which forms a uniformity with $X1$. Here, there are two possibilities. First, suppose that $X1$ is simply $X2$. Then, what you get is:

(17) a. $s_1 \models \ll \text{murderer}, X1 \gg$
b. $s_2 \models \ll \text{insane}, X1 \gg$

That is, there is an insane murderer, whoever it is (A-use). Next, suppose $X1$ and $X2$ are different roles. If $X1$ is not familiar to you, then you newly introduce it. Now what you have is that there is a murderer, there is an insane individual, and they are the same individual in reality (R-use). Note that $X1$ itself is not subject to the familiarity requirement, so it could be old or new.

Now, consider Scenario 1. Assume, for example, that Watson and Holmes have known Jones, Bond and Smith for some time. Holmes's utterance of (1) is an instance of A-use, so Watson's informational state after interpreting it can be partially described as (18). The "murderer" role *is* the "insanity" role, which is represented as x_1. Thus, murdering and insanity are grouped together in a single box. Each box constitutes a single story or a piece of information.[10]

[8] Roughly speaking, the version of Situation Theory illustrated in Devlin (1991) or Gawron & Peters (1990) is assumed (for example, in (16) s_1 is the subject's resource situation and s_2 is the whole statement's described situation). However, nothing crucial hinges on this. Also note that things like $X1$, $X2$ or N should be interpreted in the way illustrated below.

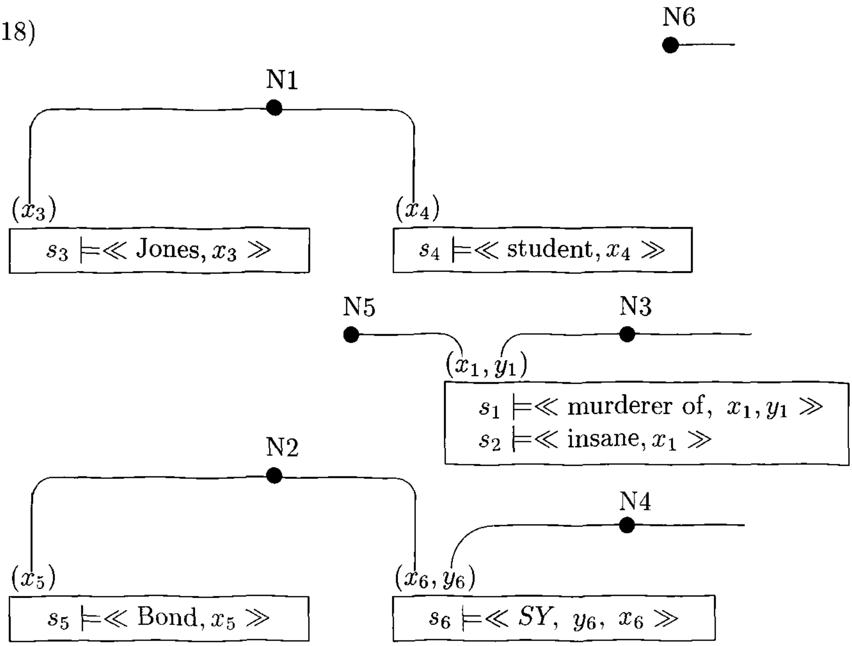
[9] Technically, one might want to formalize N as an equivalence class of roles, in which case the arrow relation can be thought of as a membership relation.

[10] In the following graphical representations, 'Jones' and 'Bond' stand for the prop-

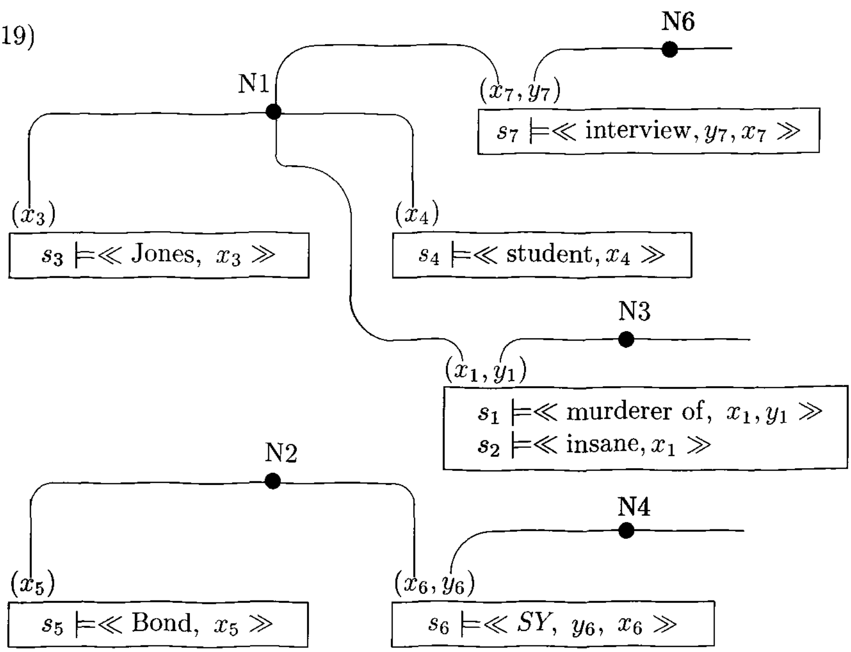
We represent uniformities as nodes. Intuitively, $N1$ is Jones, $N2$ is Bond, $N3$ is Smith, $N4$ is Watson himself, and $N6$ is Holmes. Probably Watson knows a variety of stories about these four people. Here we simplify things and write only some of them. Note that x_6 is a *role* in a visually perceived scene. Direct confrontation is also a piece of information. Watson thinks that x_5 and x_6 are one and the same person, which is represented in terms of role linking through the node $N2$. Similarly for $N1$, $N3$ and $N6$. But x_1 is not linked to any other role. That is, $N5$ is not connected with any other role.

erties of being named Jones and of being named Bond, respectively. SY stands for the relation of seeing yesterday.

(18)


N6
N1
(x_3)
(x_4)
$s_3 \models \ll \text{Jones}, x_3 \gg$
$s_4 \models \ll \text{student}, x_4 \gg$
N5
N3
(x_1, y_1)
$s_1 \models \ll \text{murderer of}, \ x_1, y_1 \gg$
$s_2 \models \ll \text{insane}, x_1 \gg$
N2
N4
(x_5)
(x_6, y_6)
$s_5 \models \ll \text{Bond}, x_5 \gg$
$s_6 \models \ll SY, \ y_6, \ x_6 \gg$

(19)

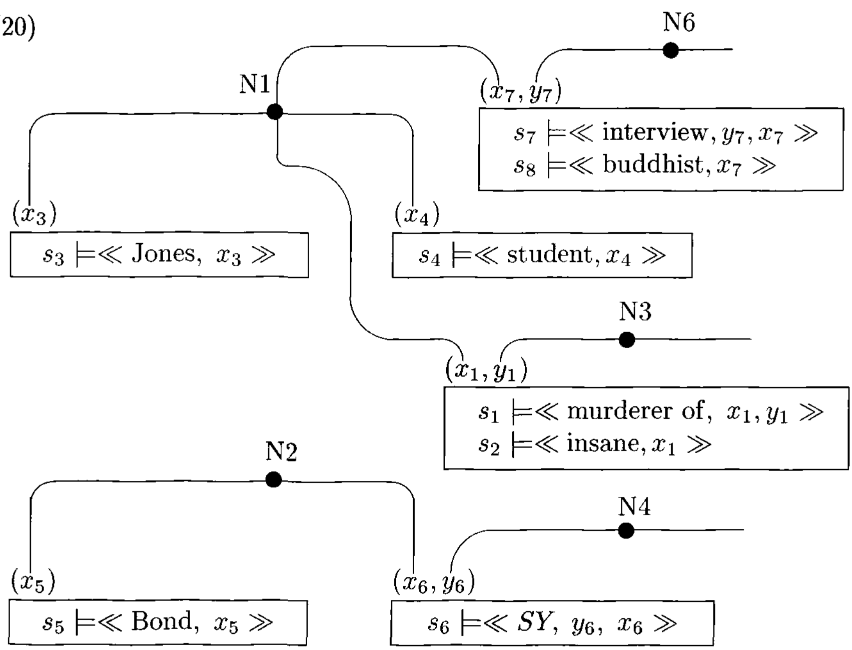

N6
N1
(x_7, y_7)
$s_7 \models \ll \text{interview}, y_7, x_7 \gg$
(x_3)
(x_4)
$s_3 \models \ll \text{Jones}, \ x_3 \gg$
$s_4 \models \ll \text{student}, x_4 \gg$
N3
(x_1, y_1)
$s_1 \models \ll \text{murderer of}, \ x_1, y_1 \gg$
$s_2 \models \ll \text{insane}, x_1 \gg$
N2
N4
(x_5)
(x_6, y_6)
$s_5 \models \ll \text{Bond}, \ x_5 \gg$
$s_6 \models \ll SY, \ y_6, \ x_6 \gg$

Next, Holmes utters (2). *Jones* picks up x_3 and *Smith's murderer* picks up x_1. Holmes's utterance requires that they share a node. Thus, $N1$ and $N5$ merge. Call the result $N1$. Further, Watson learns that Holmes interviewed somebody and that the interviewee is Jones. Then, Watson's informational state changes, as illustrated in (19). Note that x_7 and y_7 are new roles introduced by the use of the verb *interviewed.*

Thus, Watson believes that being an insane murderer is another story about Jones. Similarly for being interviewed by Holmes.

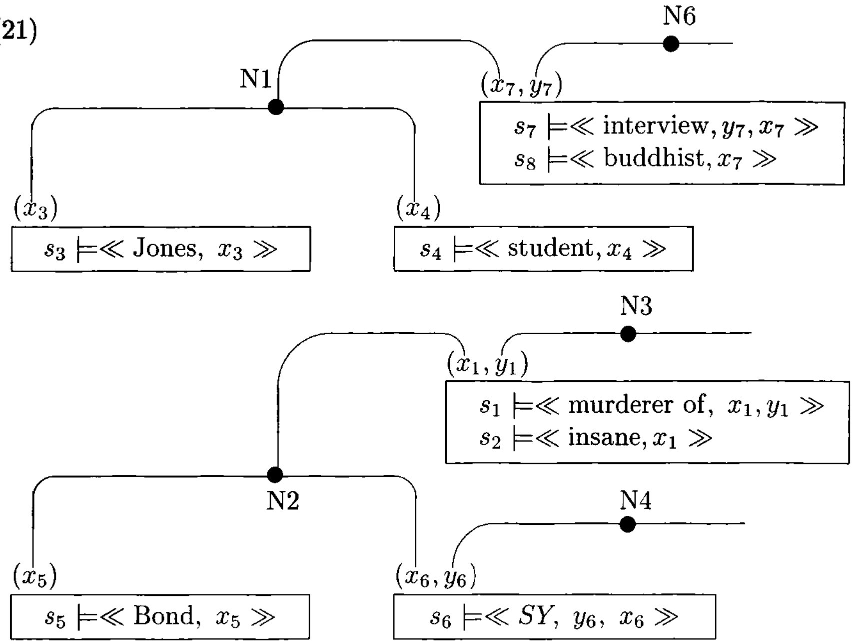
Next, Holmes utters (3). *Smith's murderer* picks up x_1, and Buddhism is predicated on x_7, which is linked to x_1 through $N1$. Thus, Watson's informational state becomes (20). Node connections are not changed, but new information on x_7 is added in the upper right box.

(20)



At stage 2, Holmes utters (4). *Bond* picks up x_5 and again, *Smith's murderer* picks up x_1. These two roles should be linked, according to Holmes, while x_1 should be disconnected from x_3, the role picked up by *Jones.* The result is something like (21). This time, only node connections are changed.

(21)



This time, being an insane murderer is a story about Bond, the node $N2$. On the other hand, being an interviewed Buddhist is still a story about Jones.

In this theory, reference is a relation between a role and other roles through a node, or more intuitively, linking of different clusters of information. Exploiting a link between two different roles, we can use an expression which picks up one of them and say something about the other. As a special case (for example, Scenario 3), we can use a name, which picks up a "being named" role (for example, x_3), and say something about another role linked to it (for example, x_1). Thus, all of the scenarios above can be explained in a uniform manner.[11] Our resulting R/A distinction based on the descriptive content conception can be summarized as in (22).

(22) Let x_1 be the role picked up by an NP contained in a statement.
A-use arises when the statement predicates something on x_1.
R-use arises when the statement predicates something on some other role linked to x_1.

[11]The difference between Scenarios 1 and 2 is that, in the latter, x_7 does not get linked to any other role except x_1.

19.2.3 Analyzing the Particular Individual Conception

Now, even when a definite description is used referentially in this sense, we are not always willing to say that the speaker has a particular individual in mind. Scenario 2 is a case in point.[12] But when are we willing to say so? I claim that it depends on the roles linked to the one picked up by the description.

First, consider the notion of *knowing who someone is*. For example, let x_1 be the role of "Smith's murderer" and x_2 be the role of "the guy over there." Then consider (23a-b).

(23) a. Who is Smith's murderer?
b. Who is the guy over there?

When we ask such questions, we are asking for information which enables us to create new links between roles. To (23a), *The guy over there* will be an appropriate answer, because it links x_1 and x_2. Similarly, to (23b), *The president of the U.S.* will be an appropriate answer because it links x_2 to the "U.S. president" role. Thus, *knowing who* itself is a unitary notion in our theory.

But we put more trust on some nodes than others. For example, we intuitively feel that a node connected only with a "murderer" role is not as trustworthy as a node connected with the role of "the man over there." We believe in Holmes's ability of criminal investigation, but we will probably put more trust on our visual perception.[13] Then, intuitively, we are willing to call a given node a particular individual if it is connected with a trustworthy role. Now, suppose a given NP picks up an untrustworthy role x. If we use that NP to predicate something on some other role linked to it through a trustworthy node, we get R-use in the particular individual conception. But if we use that NP to predicate something on x, then it becomes irrelevant whether its node is trustworthy. Then, we get A-use in the particular individual conception.

19.3 A By-product

Now let's consider how copular sentences are analyzed in our theory. Probably this is one of the most attractive by-products of our theory. We follow Pollard & Sag (1987, 1994) and assume that *be* is a so-called *raising verb*, i.e. it does not affect the predicate-argument structure directly. Rather, the predicative power comes from postcopular phrases. To illustrate the assumption in a Montagovian way, we can say that in (24),

[12] When Watson becomes a speaker to a third party hearer.

[13] Other possible factors affecting trustworthiness are how old the node is in one's informational state and how much information one has about that node. For example, probably some of us are willing to say that *Immanuel Kant* picks up a particular individual without having a direct confrontation with him.

(24) Jones is Smith's murderer.

Jones is of type e while *Smith's murderer* is of type $\langle e, t \rangle$. Then, in terms of predicate-argument structure, the main predicate in (24) is not equality, but rather the property of being Smith's murderer. Let's call this property P. Then, in line with our treatment of *Smith's murderer is insane*, the sentence (24) gives the constraints in (25).

(25) a. $r \models \ll \text{named}, x, \text{'Jones'} \gg$
b. $s \models \ll P, y \gg$
c. $r \models \ll \rightarrow, x, N \gg$
d. $s \models \ll \rightarrow, y, N \gg$

Because of the familiarity requirement, the hearer is supposed to be already familiar with (25b). Similarly for (25a). Then, the new information comes from (25c-d). This is why (24) is called an identity statement. On the other hand, in the case of (26),

(26) Jones is a Buddhist.

the main predicate is the property of being *a* Buddhist. Let's call this property Q. Then, (26) gives the constraints in (27).

(27) a. $r \models \ll \text{named}, x, \text{'Jones'} \gg$
b. $s \models \ll Q, y \gg$
c. $r \models \ll \rightarrow, x, N \gg$
d. $s \models \ll \rightarrow, y, N \gg$

This time, the postcopular NP is indefinite, so it does not require or presuppose the hearer's familiarity with (27b). So usually, being a Buddhist is new information. This is why (26) seems to be a predication statement. Similarly for postcopular phrases other than NP's. They all behave like indefinite NP's in terms of familiarity (we have already seen that the verb *interview* introduce a new "interviewer" role).

Thus, the intuitively felt difference between identification and predication follows with no additional stipulation. Thus, the copula itself is not ambiguous between identity and predication. Note that this account is made available by our role-linking conception of reference.

19.4 Conclusion

Our theory is a plausibility demonstration of the hypothesis that what we should discard is Assumption 1, not Assumption 2. The R/A distinction is not a singular/general distinction but rather best seen in terms of information grouping under an ecological realist conception of individuals as uniformities. Further, this theory gives a straightforward analysis of copular sentences in a uniform manner.[14]

References

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes.* Cambridge, MA: MIT Press.

Devlin, Keith. 1991. *Logic and Information.* Cambridge, UK: Cambridge University Press.

Donnellan, Keith S. 1966. Reference and Definite Description. *Philosophical Review* 75:281–304.

Fauconnier, Gilles. 1985. *Mental Spaces: Aspects of Meaning Construction in Natural Language.* Bradford Books. Cambridge, MA: MIT Press.

Gawron, Jean Mark, and Stanley Peters. 1990. *Anaphora and Quantification in Situation Semantics.* CSLI Lecture Notes Number 19. Stanford: CSLI Publications.

Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases.* Doctoral dissertation, University of Massachusetts, Amherst.

Ishikawa, Kiyoshi. 1995. *A Network Theory of Reference.* Doctoral dissertation, University of Tokyo.

Kaplan, David. 1989. Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals. In *Themes from Kaplan*, ed. Joseph Almog, John Perry, and Howard Wettstein. 481–614. New York: Oxford University Press.

Kripke, Saul A. 1979. Speaker's Reference and Semantic Reference. In *Contemporary Perspectives in the Philosophy of Language*, ed. Peter A. French, Theodore. E. Uehling, Jr., and Howard K. Wettstein. 6–27. Minneapolis: University of Minnesota Press.

Kronfeld, Amichai. 1986. Donnellan's Distinction and a Computational Model of Reference. In *Proceedings of the 24th Annual Meeting*, ed. Alan W. Biermann, 186–191. Association for Computational Linguistics.

Pollard, Carl, and Ivan A. Sag. 1987. *Information-based Syntax and Semantics, I: Fundamentals.* CSLI Lecture Notes Number 13. Stanford: CSLI Publications.

Pollard, Carl, and Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar.* Chicago: University of Chicago Press.

Recanati, François. 1981. On Kripke on Donnellan. In *Possibilities and Limitations of Pragmatics*, ed. Herman Parret, Mariana Sbisà, and Jef Verschueren. Studies in Language Companion Series Volume 7. Amsterdam: John Benjamins B. V.

Searle, John R. 1979. Referential and Attributive. *The Monist* 62:190–208.

[14]Other possible targets of this theory include attitude reports in general and so-called *cleft* and *pseudo-cleft* constructions. See Ishikawa (1995) for details

Shimojima, Atsushi. 1993. Definite Descriptions Revisited by Situations. Paper read at the ACL/LSA Fourth Conference on Logic and Linguistics, Columbus, Ohio, July 31–August 1.

Soames, Scott. 1986. Incomplete Definite Descriptions. *Notre Dame Journal of Formal Logic* 27:349–375.

20

Where Monsters Dwell

DAVID ISRAEL AND JOHN PERRY

Introduction

> Are there such operators as 'In some contexts, it is true that', which when prefixed to a sentence yields a truth if and only if in some context the contained *sentence* (not the content expressed by it) expresses a content that is true in the circumstances of that context?
>
> Operators like 'In some contexts it is true that', which attempt to meddle with character, I call *monsters*. I claim that none can be expressed in English...
>
> I am not saying we could not construct a language with such operators, just that English is not one. And such operators *could not be added to it.* (Kaplan 1989, pp. 510f.)

Kaplan says that monsters violate **Principle 2** of his theory. **Principle 2** is that indexicals, pure and demonstrative alike, are directly referential. In providing this explanation of there being no monsters, Kaplan feels his theory has an advantage over double-indexing theories like Kamp's or Segerberg's (or Stalnaker's), which either embrace monsters or avoid them only by ad hoc stipulation, in the sharp conceptual distinction it draws between circumstances of evaluation and contexts of utterance. We shall argue that Kaplan's prohibition is also essentially stipulative, and that it is too general. The main difference between ourselves and Kaplan is that the basic carriers of a truth-value is a sentence-in-a-context; our account is utterance-based.

Our utterance-based theory, which we call the *reflexive-referential theory*

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

differs from Kaplan's in a couple of important respects, which, we claim, are crucial for a correct understanding of issues related to monsters. Here is a summary of the similarities and differences:

- On both approaches, monsters are formally possible; that is in Kaplan's formal theory, coherent definitions can be written for monsters, and this is also true on our theory.
- Conceptually, in Kaplan's "direct reference" semantics, the prohibition against monsters has what one might call a deep semantical motivation, for the basic semantical unit, the *content*, simply does not include the parameters on which monsters would operate.
- In contrast, our basic semantic unit, which we call *reflexive content*—one example of which is *indexical content*—is one among many levels of content that we recognize. Reflexive content does include the parameters on which monsters operate, and there is no deep semantic motive for excluding them on our theory.
- On our account, there are many places where monsters might, but don't dwell, and the reasons for their absence are basically pragmatic.
  - (i) We are usually interested in what we call *incremental content*, and at this level of content the parameters monsters need to thrive are unavailable
  - (ii) Reflecting these interests, some important operators, like "says that", operate only on incremental content.
- Still, we think monsters, in particular *epistemic* or *cognitive*, as opposed to *alethic*, are possible and (who knows?) even actual.

The structure of the rest of the paper is as follows. In §2, we sketch the background of Index and Double-Index theories of modality. In §3 we give a brief account of Kaplan's framework and of his theory of indexicals. In §4, we return, with fresh motivation, to Double-Index accounts. In §5 we end with a look ahead to our own account.

20.1 Modality And Monsters

20.1.1 Index Theory

A central notion of semantics in the style of Tarski is that of the truth (satisfaction) of a sentence (formula) *in a model*. This relativity of truth can seem quite artifactual, since it may seem we are interested in truth *simpliciter*. Of course, logicians aren't interested in truth simpliciter, but in logical truth. In the model-theoretic tradition, this latter notion is captured by that of truth in all models. When we move to modal logic, the extension of the classical relativity to models, in terms of truth in a world in a model, seems actually less artifactual than the base case, for we are after logical aspects of a notion that involves relativizing truth to a space of alternative possibilities. Something similar holds for temporal logic: here it is not alternative possibilities, but simply different moments in time that we conceive truth as relative to. And so on for other modal modelings. Moreover, we may have occasion to model relativities along more than one dimension.

Consider the following progression:

- $\mathcal{M}, w \models \Box\Phi$
  - iff $\forall w' : wRw' \rightarrow w' \models \Phi$
- $\mathcal{M}, t \models H\Phi$[1]
  - iff $\forall t' : t' < t \rightarrow t' \models \Phi$
- $\mathcal{M}, \langle w, t\rangle \models \Box(H\Phi)$
  - iff $\forall w', t' : wRw' \,\&\, t' < t \rightarrow \langle w', t'\rangle \models \Phi$

Why stop with two? Indeed, the advice of a great logician tells us there is no good reason to stop at all.

> For more general situations one must not think of the $i \in I$ as anything as simple as instants of time or possible worlds. In general we have
>
> $$i = \langle w, t, p, a \ldots\rangle$$
>
> where the index i has many *coordinates*...All these coordinates can be varied, **perhaps independently**, and thus affect the truth-value of statements which have indirect reference to these coordinates. (Scott 1970)

The spirit of this advice is to give a unified treatment of modality and indexicality. In the beginning sentences are evaluated with respect to or in a model; the meanings (intensions) of sentences can be thought of as properties or sets of models: the meaning of a sentence being the set of models that make it true. When we move to the alethic modalities, sentences are evaluated relative to a model and a world of that model. Fixing a model, then, we identify the meaning of a sentence with the set of worlds of that model at which the sentence is true.

Imagine one starts, where our little progression of relativities ends, with a language with both alethic and temporal modalities, but without indexicals. The meaning of a sentence Φ, $[\![\Phi]\!]$, is a set of ordered pairs of worlds and times (or a function from such pairs to $\mathcal{B} = \mathbf{2}$). If we consider adding indexicals such as the personal pronoun 'I' and the locative adverb 'here' to the language, Scott advises us simply to extend the structure of the indices

[1]'H' is the Priorean 'throughout history' or 'always in the past' operator.

or *points of reference*. The meanings of sentences of the new language are now sets of quadruples of worlds, times, individuals (people) and places. These quadruples are the *circumstances* within (against) which a sentence must be evaluated for truth and falsity.

Kaplan presents a dilemma for this *pure index theory* approach. Consider the two sentences:

(1) a. I am here now.
b. John Perry is in Moraga on June 15, 1994.

If we consider the quadruple i that consists of the actual world, June 15, 1994, JRP and Moraga and consider $[\![\text{I am here now}]\!](i)$, we can see that the proposition expressed by (1a) at that index is the same as the proposition expressed by (1b) at that index and at many others. Indeed, we can imagine or stipulate that the only relativity in (1b) is pure alethic relativity; its truth depends only on what world is being considered; it is true in some, false in others. What of (1a)? Consider the index i' that consists of the actual world, June 15, 1994, Napoleon Bonaparte and Moraga. Surely $[\![\text{I am here now.}]\!](i') = 0$. Thus (1a), too, is contingent—true at some indices, false at others. Nothing (so far) can be said against this index, for Scott explicitly allows the possibility of of independent variation of the coordinates. So unless something more is said, we seem to have lost any chance of a logic of indexicals; for, of course, (1a) differs from (1b) precisely in being a very plausible candidate for a valid sentence in such a logic.

This problem of missed validities motivates the move to restriction to *proper* indices. In our example, proper indices are those $\langle w, t, a, p\rangle$ such that in w, a is located at p at t. Now if we restrict our structures to those *proper structures* in which all indices are proper, (1a) comes out logically true: true at every index in every proper structure. Consider (2)

(2) Necessarily, I am here now.

$\Box\Phi$ is true at an index in a structure if Φ is true at every index in that structure. Now if we assume the principle of modal generalization that if $\models \Phi$, then $\models \Box\Phi$, we seem to be stuck with the logical truth of (2), and this seems wrong. Here we haven't missed a validity; we've created spurious ones.

There is a way around this dilemma. It involves dropping the standard principle of modal generalization, and Montague (arguably) avails himself of it in Montague 1974. There he allows structures with improper indices, but defines logical truth as truth at every proper index in every structure—thereby guaranteeing the logical truth of (1a). As for (2), it is not logically true, because there is a structure with improper indices, that is, a structure such that (1a), though logically true, is not (just plain) true at every index in that structure.

This solution seems to be merely a technical trick; as we shall see, though, it is an attempt to get at something real and important.

20.1.2 Double-Index Theory

What is needed, but not provided by index theory, is an explanation of the special role of proper indices in the characterization of the logical truths of a language with indexicals. This special role is determined by the different roles played by aspects of context and aspects of circumstance in determining the truth of sentences. The aspects of context—the identity of the speaker, the place and time of the utterance—determine aspects of the proposition expressed by the sentence in the context; the proposition so determined is then to be evaluated for truth and falsity in varying circumstances. We thus get a two-step account:

- from sentences and contexts to propositions,
- from propositions in circumstances to truth-values.

Index theory yields a one-step account: from sentences at indices to truth-values. Kaplan's diagnosis of the problems with index theory is thus exactly right:

> The difficulty is the attempt to assimilate the role of *context* to that of *circumstance*. The indices $\langle w, t, a, p\rangle$ that represent contexts must be proper in order that (1a) be a truth of the logic of indexicals, but the indices that represent circumstances must include improper ones in that (2) *not* be a logical truth.
> If one wishes to stay with this sort of index theory, the minimal requirement is a system of *double* indexing, one index for context and another for circumstances. (Kaplan 1989, pp. 509f.)

Double-Index theory was developed by Vlach and Kamp (Kamp 1971), both students of Montague and systematized, though not with complete generality, by Segerberg (Segerberg 1973). We present here only a hint by way of examples. The basic idea is to model the distinction between contexts and circumstances via a two-dimensional logic of (one family) of index sets. In the simplest case, the index set is a set of points, as in abstract versions of modal logic. We introduce the following notation:

- $w \models_v \Phi$

which, for the applications at hand, is to be read "the sentence Φ, uttered in world w is true at world v. Consider now the following contrasting pairs:

(3) a. Necessarily, Φ
- $w \models_v \Box\Phi$ iff $\forall v' : vRv' \rightarrow w \models_{v'} \Phi$

b. Actually, Φ
- $w \models_v A\Phi$ iff $w \models_w \Phi$

(4) a. On the next day, Φ

- $t \models_u O\Phi$ iff $t \models_{u+1} \Phi$

b. Tomorrow, Φ

- $t \models_u T\Phi$ iff $t \models_{t+1} \Phi$

How does double-indexing deal with (1) and (2)? Rather than look at that rather complicated case, it suffices to look to see how double-indexing allows one to guarantee the validity of $\Phi \leftrightarrow A\Phi$ without yielding that of $\Box(\Phi \leftrightarrow A\Phi)$. The idea is precisely that deployed in the single index account by Montague. Let indices now be pairs of worlds and call the diagonal pairs $\langle w, w\rangle$ proper. Now define indexical validity or indexical logical truth as truth at all proper indices in all structures. Validity or logical truth simpliciter is truth at all indices. Thus the truth-clause for A guarantees the indexical validity of $\Phi \leftrightarrow A\Phi$, but necessity requires truth at all indices, proper or not.[2] Thus for the necessitation to be indexically valid, the original biconditional must be true at all indices, proper or not; but we can have $w \models_{w'} A\Phi$ without having $w \models_{w'} \Phi$.

Kaplan does not take note of the device Montague exploited to solve the problem for index theory posed by (1) and (2); nor does he mention the similar device available to double index theory. His complaint against the latter is that it, too, blurs the distinction between contexts and circumstances. This seems false, but it does lead Kaplan into his discussion of monsters.

> However, mere double indexing, without a clear conceptual understanding of what each index stands for [of the conceptual difference between context and circumstance?] is still not enough to avoid pitfalls. (Kaplan 1989, p. 510)

The pitfall is the begetting of monsters. What are monsters? Kaplan's example is:

- In some context it is true that Φ

This is a monster if it understood as yielding a truth upon being prefixed to a sentence just in case in some context, *not* in some circumstance, the embedded sentence expresses a true proposition in the circumstances associated with that context. Kaplan notes that there is a construction in English that allows us to say what we seem to want to say with this monster:

- In some context, "Φ" is true.

or more fully, and in the style of double-indexing:

[2] We assume, for simplicity, that necessity is a universal, S-5 operator.

- In some context, c, the proposition expressed by "Φ" in c is true in c.

This semantic ascent brings out a parallel between monsters and a famous brand of puzzle:

- If you call a horse's tail a leg, how many legs does a horse have?

On a monstrous reading, the answer is: 1. This is the reading—assumed here to exist—on which the puzzle is paraphrasable as:

- In a context in which the word 'leg' means what 'tail' actually means, how many legs does a horse have?

Now we can see the aptness of Kaplan's comment that monsters attempt to operate on the meanings of sentences, as opposed to the content (proposition) expressed by a sentence in a context. Of course there is a such a difference only in an account in which the meaning of a sentence differs from the proposition expressed by it. There is no such difference on pure index theory. The meaning of sentence (in a structure), $[\![\Phi]\!]$ is that function from indices of that structure to $\{0, 1\} = \mathcal{B}$ whose value for $\imath \in I$ is 1 just in case Φ is true at $\imath$, and this is what is usually taken to be the proposition expressed by Φ.

In Double Index Theory, the meanings of sentences are functions from indices (in their role as contexts) to propositions, which are themselves functions from indices (in their role of circumstances of evaluation) to truth-values:

- Meanings of sentences as functions:
- $\mathcal{I} \rightarrow (\mathcal{I} \rightarrow \mathcal{B})$

A two-dimensional operator corresponds to a function $\mathcal{F}$: $(I \rightarrow (I \rightarrow B)) \rightarrow (I \rightarrow (I \rightarrow B))$. **All such operators are monsters!** Of course, some of these monsters are benign, in that they don't really change meanings: they operate only on proposition expressed. Some, however, are not so innocent.

To make this contrast more intelligible, we first introduce some notational conventions.

- $[\![\Phi]\!](i)(i') = 1$ iff $i \models_{\imath'} \Phi$.

We want to be able to get at the proposition expressed by a sentence at an index. To do this, we superscript the context-index:

- $[\![\Phi]\!]^{\imath}(i') = [\![\Phi]\!](\imath)(i')$

Now let $[\![\oslash\Phi]\!]^{\imath}$ be the proposition such that $[\![\oslash\Phi]\!]^{\imath}(i') = [\![\Phi]\!](i)(i)$. Thus, for any index (= context) i, $[\![\oslash\Phi]\!]^{\imath}$ represents the diagonal proposition for i, so $[\![\oslash\Phi]\!]$ is the 'diagonalizing' of the meaning $[\![\Phi]\!]$ of Φ. Now consider the following two operators, the first due to Kamp, the second to Vlach:

- $[\![\ddagger\Phi]\!](i)(i') = [\![\oslash\Phi]\!]^{i} = [\![\Phi]\!](i)(i)$
  - Doesn't change the context parameter
- $[\![\dagger\Phi]\!](i)(i') = [\![\oslash\Phi]\!]^{i'} = [\![\Phi]\!](i')(i')$

This last comes to: the proposition expressed by Φ is true; where "the proposition expressed by Φ" is *non-rigid*, varying with circumstance of evaluation. This is a nonbenign monster.

Kaplan's point is just that Double Index Theory makes no principled distinction between these two operators:

> [Double Index Theory] allows a simple and elegant introduction of many operators which are monsters. In abstracting from the distinct conceptual roles of played by contexts of use and circumstances of evaluation the special logic of indexicals has been obscured. Of course restrictions can be put on the two-dimensional logic to exorcise the monsters, but to do so would be to give up the mathematical advantages of that formulation. (Kaplan 1989, p. 512)

This is as good a place as any to enter a caveat about identifying Montague as a pure index theorist. Montague distinguishes within what we might call a 'generalized index' two parts, an index proper—not to be confused with a proper index—and a context of use. The *meanings* of closed sentences are functions from generalized indices to $\mathcal{B}$; but the *senses* of sentences are functions only from indices proper to $\mathcal{B}$. This means that the analogue of two-dimensional functions are ruled out: functions from contexts (or from generalized indices) into **2** cannot be the senses (contents) of sentences and hence can not be arguments to modal (intensional) operators. No monsters!! Thus Montague's theory, like double index theory, does distinguish between meanings and contents (propositions/senses) and it prohibits monsters–at least monsters of the type Kaplan discusses. Moreover as we shall see this prohibition of monsters does not differ in form all that much from Kaplan's.

20.2 Kaplan's Theory

We have noted that a central idea in model-theoretic semantics is that of the relativity of truth. In general what a semantic account provides is a type of truth-valuable entities and a circumstance of evaluation. In the case of modal logic and its generalization in pure index theory, the truth-valuable entities are sentences and the circumstances of evaluation are indices (within structures). In the case of double index theory, the truth-valuable entities are propositions (functions from indices to $\mathcal{B}$) and the circumstances of evaluation are again indices. Given what Kaplan says in the descriptive, philosophical sections of Kaplan 1989, one might expect something similar in his logic of demonstratives, except that a clear conceptual distinction would be made between contexts and circumstances. That is, one would expect a two-step theory, modeling the meanings—Kaplan's term is 'character'—of sentences as functions from contexts to propositions, which are in turn functions from circumstances to $\mathcal{B}$. In fact, though, he present a version of a single index theory that, like Montague's, divides the one vector into two parts: context and circumstance, context itself being modeled as a quadruple consisting of an agent, a time, a position, and a world. Again, as in Montague, circumstances (indices proper) are pairs of worlds and times.[3] So the truth-evaluable entity in Kaplan's account are sentences-in-a-context:

> The Content of a sentence in a context is, roughly, the proposition the sentence would express if uttered in that context. This description is not quite accurate. First, it is important to distinguish an *utterance* from a *sentence-in-a-context.* The former notion is from the theory of speech acts, the latter from semantics. Utterances take time, and utterances of distinct sentences cannot be simultaneous (i.e., in the same context).[4] But to develop a logic of demonstratives it seems most natural to be able to evaluate several premises and a conclusion all in the same context. The notion of Φ *being true in* c does not require an utterance of Φ. In particular, c_A (the agent of the context) need not be uttering Φ in c_W at c_T. (Kaplan 1989, p. 546)

We remind the reader of Kaplan's two basic principles about demonstratives and indexicals:

Principle 1 The referent of a pure indexical depends on the context, and the referent of a demonstrative depends on the associated demonstration.

Principle 2 Indexicals, pure and demonstrative alike, are directly referential.

What Kaplan means by Principle 2 is that the referential relation between, e.g., an indexical, as occurring in a sentence Φ in a context and its referent is not mediated by the content of the sentence in that context. We return to this below.

We present a brief sketch of Kaplan's *Logic of Demonstratives* in outline form:

The Formal System $\mathcal{LD}$:

- $\mathcal{A}$ is a $\mathcal{LD}$ structure iff there are $\mathcal{C}, \mathcal{W}, \mathcal{U}, \mathcal{P}, \mathcal{T}, \mathcal{I}$ such that:
  - $\mathcal{A} = \langle \mathcal{C}, \mathcal{W}, \mathcal{U}, \mathcal{P}, \mathcal{T}, \mathcal{I} \rangle$

[3] We are, of course, ignoring relativity to assignments.

[4] This is gratuitous; one can chose the granularity of the temporal dimension to suit one's purposes.

- $\mathcal{C}$ is a nonempty set (of contexts)
- If $c \in \mathcal{C}$, then
  - $c_A \in \mathcal{U}$ (the *agent* of c)
  - $c_T \in \mathcal{T}$ (the *time* of c)
  - $c_P \in \mathcal{P}$ (the *place* of c)
  - $c_W \in \mathcal{W}$ (the *world* of c)
- $\mathcal{W}$ is a nonempty set (of worlds)
- $\mathcal{U}$ is a nonempty set (of all—actual and possible—individuals)
- $\mathcal{P}$ is a nonempty set (of positions—common to all worlds)
- $\mathcal{T}$ is the set of integers (thought of as times, common to all worlds)
- $\mathcal{I}$ is the interpretation function, assigning pairs of times and worlds to wffs., and meeting the following conditions:
- $i \in \mathcal{U}$ iff $(\exists t \in \mathcal{T})(\exists w \in \mathcal{W})(\langle i\rangle \in \mathcal{I}_{\text{Exist}}(t,w))$
- If $c \in \mathcal{C}$, then $\langle c_A, c_P\rangle \in \mathcal{I}_{\text{Located}}(c_T, c_W)$
- If $\langle i, p\rangle \in \mathcal{I}_{\text{Located}}(t,w)$ then
  $\langle i\rangle \in \mathcal{I}_{\text{Exist}}(t,w)$

* **So all contexts are proper.**
* **Truth and Content:**

Truth $\models^{\mathcal{A}}_{ctw} \Phi$ for: Φ, in context c is true with respect to time t and world w.

Denotation $|\alpha|_{cftw}$ for *the denotation of* α, in context c (under f) with respect to time t and world w.

Content Where Φ is a wff.,$\{\Phi\}^{\mathcal{A}}_c$ for the **content** of Φ in c.

- $\{\Phi\}^{\mathcal{A}}_c(t,w) =$ TRUTH iff $\models^{\mathcal{A}}_{ctw} \Phi$.

Truth in a context Φ is *true in* c in $\mathcal{A}$ iff $\{\Phi\}^{\mathcal{A}}_c(c_T, c_W) =$ TRUTH.

Validity Φ is *valid in* $\mathcal{LD}$ iff for every $\mathcal{LD}$ structure $\mathcal{A}$ and every $c \in \mathcal{A}$, Φ is true in $\mathcal{A}$.

Character $\{\Phi\}^{\mathcal{A}}(c) = \{\Phi\}^{\mathcal{A}}_c$

* **The crucial clauses of the definition of satisfaction**

1. $\models_{cftw} R\alpha_1 \ldots \alpha_n$ iff $\langle |\alpha_1|_{cftw}, \ldots, |\alpha_n|_{cftw}\rangle \in \mathcal{I}_R(t,w)$
2. $\models_{ctw} \Box\Phi$ iff $\forall w' \in \mathcal{W} : \models_{ctw'} \Phi$
3. $\models_{ctw} A\Phi$ iff $\models_{ctc_W} \Phi$
4. $\models_{ctw} N\Phi$ iff $\models_{cc_Tw} \Phi$
5. $|I|_{ctw} = c_A$
6. $|\text{Here}|_{ctw} = c_P$

* **Crucial cases**

1. $\models (\Phi \leftrightarrow AN\Phi)$
2. $\models N(\text{Located, I, Here})$

3. $\models$ Exist I
4. $\not\models \Box(\Phi \leftrightarrow AN\Phi)$
5. $\not\models \Box N$(Located, I, Here)
6. $\not\models \Box$Exist I

Now what of monsters? In Kaplan's theory, the meaning or character of a sentence Φ is a function: $\mathcal{C} \to (\mathcal{I} \to \mathcal{B})$ where $\mathcal{I} = (\mathcal{T} \times \mathcal{W})$. To rule out monsters is to rule that there are no functions of the type: **Character $\to$ Character**. This is to stipulate that no operator can effect (any component of) the c component of the index $\langle cftw \rangle$. This rules out the analogue of the † operator above and is directly analogous to Montague's prohibition of operators on contexts or generalized indices.

20.3 The Veil of Ignorance

As Kaplan makes clear, the notion of Content, of what is said, is central to his account. At the level of sentential content, of proposition expressed, the content/reference determining features of the context have been applied and as, in function application generally, no trace of them remains. The features of the sentence-in-a-context that determine what is said are not part of what is said. Operators such as the alethic and temporal modalities only apply to propositions expressed: this is precisely what the stipulation against monsters comes to. But what of other attitudes—in particular what of the propositional-attitude operators, more narrowly construed. Though Kaplan has much to say about issues concerning the interaction between indexicals/demonstratives and such operators, he does not introduce them into $\mathcal{LD}$.

Consider knowledge. Imagine we decide to add a family of unary modal operators K_a, indexed by agents $a \in \mathcal{U}$ to $\mathcal{LD}$ and to introduce (structurally identical) associated binary accessibility relations for them, where the intuitive reading of the relationship is that wK_aw' iff w' is an epistemic alternative for a relative to w. But now we should remind ourselves of the following facts about actual utterances and the contexts in which they are produced:

- One might not know who the agent of c is.
- One might not know when the time of c is.
- One might not know what the place of c is.
- One might not know what the world of c is.

Indeed, a speaker himself might be ignorant of the fact that he was the speaker of a given utterance. Consider the case of echoes, especially as produced at a famous and much-visited location. So given a type for an utterance, that is, given a sentence Φ, other contexts for Φ are epistemic alternatives. To see what may be involved here, let us return to a simple double index account, in which the basic indices are just worlds. Here the 'context' index would represent the epistemic perspective of the agent and the circumstance index would, as usual represent, the world about which the knowledge claims are made. The clause for K (now suppressing indexing by agent) would be as follows:

- $w \models_v K\Phi$ iff $\forall w', v' : \langle w, v\rangle R_K \langle w', v'\rangle \rightarrow w' \models_{v'} \Phi$.[5]

Notice that this operator involves quantification over contexts or generalized indices. It is a non-benign monster.

What would this look like in $\mathcal{LD}$? In conformity with Kaplan's restriction, and supposing for simplicity that indices proper—circumstances—are just worlds, all he would allow us is this:

- $cw \models K\Phi$ iff $\forall w' : wR_K w' \rightarrow cw' \models \Phi$.

But to capture the facts about ignorance, what we need is rather more like this:

- $cw \models K\Phi$ iff $\forall c', w' : \langle c, w\rangle R_K \langle c', w'\rangle \rightarrow c'w' \models \Phi$; where $c' = \langle c'_a, c'_t, c'_p, c'_w\rangle$.

This, of course, is monstrous.

In sum:

- Perhaps there is something right about Kaplan's prohibition, but it is not quite right. Perhaps there could not be pure modal monsters, but there can be epistemic (and deontic, etc.) monsters.
- Double indexing has no explanation of the lack of modal monsters; Kaplan's theory does not allow the epistemic ones.

We claim that an utterance-based theory explains why there can be epistemic monsters, but no modal monsters, and also clarifies Kaplan's fundamental distinction between contexts and circumstances of evaluation.

20.4 Utterances

20.4.1 The Reflexive-Referential Theory: A Look Ahead

Utterances are the fundamental truth-evaluable entities. Utterances are acts, concrete nonrepeatable particular events.[6] Utterances are not to be confused with tokens. Tokens are also concrete particulars, but they are objects, not events. Tokens are reusable, in way that utterances are not. The distinction between utterance and token is fairly easy to see with respect to written tokens. Written tokens typically have longer duration than the act—the utterance—that produced them. They are composed of chalk

[5]See Rabinowicz and Segerberg 1994 for a similar treatment of knowledge in a two-dimensional context, motivated by very different concerns.

[6]For more on acts and actions, see Israel et al. 1991, Israel et al. 1993.

or graphite, they can be erased or underlined Not so the utterance In speaking we produce more-or-less evanescent tokens whose perceptible existence doesn't much outlive the duration of the utterance that produced them Still these tokens can be recorded and then manipulated in various ways Utterances, speakings, can be recorded, too, indeed they can be filmed without sound For many years, that's what movies largely consisted in recordings of utterances, without any recording of the tokens produced In the case of computer files, the distance between utterance and token is quite large, indeed, it is a little mysterious as to what the token produced is It is utterances, not tokens, that are the primary bearers of truth and falsity

As Kaplan notes, utterances take time An utterance also has a particular speaker and a place All actual utterances have the same world, namely the actual world We claim, without argument, that the agent, time and place of an act are metaphysically essential features of that act It makes no sense to say that its world is also metaphysically essential to it

Consider now what a competent speaker/hearer of English knows about the truth-conditions of an utterance of "I am tired" solely on the basis of his linguistic knowledge, that is, in the absence of knowing who said it and when

- A utterance u of "I am tired" is true iff *the speaker of u is tired at the time of u*

The italicized condition yields a proposition when predicated of a particular utterance So consider a particular utterance **u**, it is true iff *the speaker of* **u** *is tired at the time of* **u** We call this proposition the reflexive content of **u**—more particularly, its *indexical content*[7] What Kaplan calls the content, what is said by the utterance, we call the *incremental content*, it is generated from the reflexive content, given all the features of the utterance and context that determine reference In the case of pure indexicals, these are features of the utterance itself In the case of demonstratives, these will include features of the wider context of utterance

As we have seen, the incremental content does not (usually) involve the utterance or its context, these have been used in determining what is said, they are not a part of it But that need not be true for the other contents[8] Assume a competent speaker/hearer of English encounters **u**, but is ignorant of who said it and when He knows its truth-condition, its pure reflexive content In this case, he is ignorant of its content in two

[7] This is precisely to leave open the possibility of there being other kinds of reflexive content, e g , that associated with uses of proper names

[8] Note that in general where there are n independent dimensions of context-relativity exploited in an utterance, there will be $2^n - 1$ reflexive contents

different ways: he is ignorant of what the content is, of what proposition the utterance expresses, because he is ignorant of the context of utterance, and he is ignorant of the truth-value of that proposition. Formalizing such facts motivates the monstrous treatment of knowledge sketched above.

20.4.2 Monsters, Revisited

We have claimed that agent, time and place of an utterance are essential features of it. This explains why there cannot be metaphysical monsters. You can't 'take' an utterance to a metaphysical alternative and leave its reference and truth determining features behind. That is, there are no real modal alternatives with respect to the context of utterance.

One can, on the other hand, take a sentence uttered in one index to another index, and one can take a sentence and an agent to another world and another time. Thus if, like Kaplan, one takes as the prime truth-evaluable entity a sentence-in-a-context, where a context is an n-tuple but without a representative of the particular utterance, it is hard to justify the stipulation that no expression of the language can involve a shift in context. Indeed we have seen that because none of the metaphysically essential features of utterances are epistemically transparent features, in modeling knowledge it seems that we want to be able to shift context. We should note that the claim of asymmetry between modal and epistemic monsters does not commit us to the view that there are real monsters in English.

References

Israel, David, John Perry, and Syun Tutiya. 1991. Actions and Movements. In *Proceedings of IJCAI'91*. Sydney. Morgan Kaufmann.

Israel, David, John Perry, and Syun Tutiya. 1993. Executions, Motivations, and Accomplishments. *Philosophical Review* 102:515–540.

Kamp, Hans. 1971. Formal Properties of 'Now'. *Theoria* 37:227–274.

Kaplan, David. 1989. Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals. In *Themes from Kaplan*, ed. Joseph Almog, John Perry, and Howard Wettstein. 481–614. New York: Oxford University Press.

Montague, Richard. 1974. Pragmatics. In *Formal Philosophy*, ed. Richmond Thomason. 95–118. New Haven: Yale University Press.

Rabinowicz, Wlodek, and Krister Segerberg. 1994. Actual Truth, Possible Knowledge. In *Proceedings of TARK V*, ed. Ron Fagin, 122–137. Asilomar. Morgan Kaufmann.

Scott, Dana. 1970. Advice on Modal Logic. In *Philosophical Problems in Logic*, ed. Karel Lambert. 143–173. Dordrecht: Reidel.

Segerberg, Krister. 1973. Two-Dimensional Modal Logic. *Journal of Philosophical Logic* 2:77–96.

21

A Distributed System Model for Actions of Situated Agents

YASUHIRO KATAGIRI

Introduction

Knowledge, beliefs, goals, intentions, these mental states of an agent partially determine what actions are performed by the agent, and actions in turn cause changes in mental states of the agent. Developing a model in which to express and investigate relationships between agent's mental states and actions is essential both to the theory of agents and to the design of artificial agents.

Several attempts has been made to develop a model to investigate these relationships (Moore 1985, Cohen and Levesque 1990, Levesque et al. 1990, Perrault 1990, Rao and Georgeff 1991). Most of these models rely in some form or other on modal logical formalisms. Modal logical formalisms are powerful in the sense that they provide us with a way to state precisely what conditions on both mental states and actions have to obtain to successfully attain certain states, thereby giving us a way to judge what inferences are legitimate about the courses of events brought about by acting agents.

A question we should raise here is who this *"us"* should be. A standard, commonly accepted and somewhat cautious answer seems to be theorists/designers. We could regard modal logical formulations as specifications of the behavior of agents from theorists/designers' perspective. On the other hand, inference about courses of events brought about by acting agents has to occupy a central part in action planning, reasoning performed by agents themselves, and hence it would be nice if *"us"* also includes agents, and the application of the formalism could extend to planning domains.

However, modal logical specifications of mental states and actions do not

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

provide sufficiently detailed information to be useful for a situated agent in making inferences on the effects of her own actions. Why it is so can be seen from the following three observations. First, modal logical specifications of mental states really state conditions on the relationships between an agent and her surrounding environment. Possible world semantics of modal logics of knowledge and belief interprets them in terms of the agent's epistemic (or doxastic) alternatives, what the world or courses of events can possibly be like given the epistemic (or doxastic) states of the agent. Since the range of alternative world states or alternative courses of events depend on what environment the agent is located in, modal logical specification of the agent's mental states actually is not a specification of the agent's states per se, but a specification of conditions on the relationship between the agent's internal states and the environmental states. Secondly, an agent has to decide on what action to take based solely on her internal states. All conditions on external environmental states must first be reflected, through perception and reasoning, in her internal states before they can play any role in her action selection. Thirdly, because of her situated nature, an agent can have only partial information on her surrounding environment. Since situated nature of agents prevents her from having complete information on her surrounding environment, she cannot have a complete idea by herself what kind of environment she is placed in. These three observations together indicate that modal logical specifications are not sufficient for a situated agent to reason about and to decide on her own actions. Even though modal logical specifications are sufficient for theorists/designers, who have the complete grasp of the environment, to reason about an agent's actions, the agent herself cannot reason about her own actions because she doesn't know what her epistemic/doxastic alternatives really are.

To be of some use to situated agents reasoning and acting on partial information, it would be better to have a formalism where internal states of agents and states of the environment can independently be expressed and conditions on action success can be stated as conditions on relationships that must hold between them. Recent developments in distributed computing research has revealed (Halpern and Moses 1990) that a knowledge-based paradigm for analyzing distributed systems can provide us with a simple yet powerful method for analyzing behaviors of programs interacting both with each other and with their environment. Since the analysis refers both to programs and to their knowledge, it is expected that this paradigm could be extended to analyzing actions of situated agents which is also useful to agents themselves.

We investigate, in this paper, a distributed system framework in which to reason about actions performed by situated agents. We will first give an analysis of knowledge and actions in terms of distributed systems, in which we emphasize the importance of distinguishing environment-dependent notions of knowledge and actions from environment-independent notions of judgments and acts. We will then state action success conditions in terms of implementation conditions between acts and actions. We then demonstrate the effectiveness of the framework through the analysis of search examples, where we show that `INFORM` action can be conceived of an example of a test action in a joint action setting.

21.1 A Distributed System Model of Knowledge and Action

A distributed system model assumes one or more agents executing their programs in an environment. Execution of programs can bring about changes both in environment and in agents. Environment changes correspond to physical actions, whereas agent changes correspond to informational actions, such as perceiving or sending/receiving of messages. It is not guaranteed that execution of programs brings about a constant effect. One agent's program step may interfere with other agents' program step, and environment may also non-deterministically intervene program execution and change the outcome. By regarding the environment as executing a non-deterministic program, the entire system is modeled as a large transition system.

Three characteristics are worth noting for a distributed system model as a model of situated agent actions. First, by including the environment as a separate and independent component similar in status to other agents, the model provides us with a way to give explicit descriptions of the interaction of agents and the environment. Secondly, since programs run on agents' internal states only without direct access to states of the environment, the model captures the partial information availability for situated agents. Finally, the model gives us a unified framework for both multi-agent interaction and agent-environment interaction.

21.1.1 Distributed Systems

A distributed system consists of n agents $\{a_1, \ldots, a_n\}$ and an environment e. A local state s_i of each agent a_i at a given instant is taken from a corresponding set of local states S^i. A local state s_e of the environment e is similarly taken from a set S^e. A global state s of the system is thus a tuple consisting of $n+1$ local states, $\langle s_e, s_1, \ldots, s_n \rangle$, and the behavior of the system can be specified as transitions between global states. For expository convenience, we regard the environment as the 0-th agent and write a_0 and s_0 for e and s_e, respectively.

A distributed system M is a tuple $\langle A, S, BACT, \Pi, \tau \rangle$, where A is a finite set consisting of agents and the environment. S is a set of global states, that is, the product of sets of local states S^i for all $a_i^{\prime}s$. ACT is a tuple consisting of sets of acts $BACT_i^{\prime}s$, each of which corresponds to the range of basic acts that can be executed by each a_i. Π is a *protocol* which specifies actual programs, that is, which act act_i to execute when, for each agent and the environment.

$$act_i \in \Pi(a_i, s_i) \subseteq BACT_i$$

We assume that the environment may act non-deterministically, but for other agents programs are deterministic and Π uniquely determines act_i. τ is a *transition function* which specifies the transition of the system given all the acts executed by agents and the environment in a global state.

$$s_{t+1} = \tau(act_e, act_1, \ldots, act_n)(s_t)$$

We simply assume that a proposition corresponds to a set of global states. The notion of a proposition holding in a state reduces to set membership. We denote a set of global states in which a proposition p holds by S_p

21.1.2 Knowledge

It has been shown (Halpern and Moses 1990) that S5 modal logical interpretation of knowledge can quite naturally be applied to the distributed system paradigm by first taking each global state as a possible world and then taking global states which share the same local state of an agent as epistemic alternatives with each other for the agent. Since the agent has no way of distinguishing among global states, if they correspond to the same local state, propositions holding in all of those states must be known by the agent.

Definition 1 [Knowledge] We say a_i *knows* p in s with respect to U, and write $K_{a_i}^U p$ iff $\forall s' \in U.(s_i' = s_i) \supset (s' \in S_p)$

As has been noted in the previous section, this conception takes knowledge as the correspondence between the agent's internal states and the environmental states, which is only accessible to theorists/designers. What an agent knows depends on what kind of environment the agent is interacting with. In order to obtain an analysis of knowledge useful to agents, we have to fix a program and vary the environment.

For each proposition p, we assume there is a corresponding division of local states S^i of an agent a_i into three parts, s_p^+, s_p^-, and $\mathrm{s}_p^?$. They are supposed to mean "a_i knows that p," "a_i knows that not p," and "a_i does not know whether p or not," respectively. But there is no guarantee that the division actually corresponds to the knowledge or the lack of knowledge of p under various environments. No matter what the actual correspondence, the division of the local states amounts to the *judgment* made by the agent

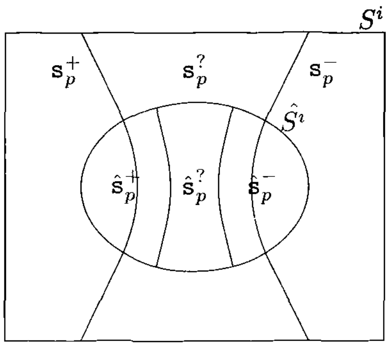


FIGURE 1 Division of agent's internal states.

on p, and it is referred to in the agent program Π in determining what act to perform next.

Definition 2 [Judgment] A *judgment* is an internal test of the form $s_i \in s_p^{+/-}$ performed in agent a_i's programs to categorize its local states.

Imagine an agent program is executed in a certain environment starting from a certain initial state. Because of non-deterministic nature of the behavior of the environment, to each time point in the execution there corresponds a set of global states, which in turn determines the correct division of the local states concerning a proposition p *at that time point in that environment.* We denote by $\hat{S}$ the range of global states that is possible at the point in execution, and by $\hat{S}^i$ its projection to the set of local states S^i of the agent a_i. We further denote by $\hat{s}_p^+$, $\hat{s}_p^-$ and $\hat{s}_p^?$, the correct division of $\hat{S}^i$ into the three knowledge states at that execution point in that environment. The division of an agent's internal state in terms of knowledge and judgment on p is depicted in Figure 1.

Matching between s_p^+, s_p^-, $s_p^?$, and $\hat{s}_p^+$, $\hat{s}_p^-$, $\hat{s}_p^?$ indicates how adequate a judgment the agent can make about her environment. We can distinguish three conditions that are significant in evaluating this adequacy.

- **Accuracy**: all judgment is correct.
  $(s_p^{+/-} \cap \hat{S}^i) \subseteq \hat{s}_p^{+/-}$
- **Maximality**: all judgment is maximal in the sense that all knowledge is guaranteed to be incorporated in the judgment.
  $(s_p^{+/-} \cap \hat{S}^i) - \hat{s}_p^{+/-} = \phi$
- **Information availability**: information on whether p is completely available.
  $\hat{s}_p^? = \phi$

Accuracy or maximality conditions may hold for either positive or negative judgments alone. Accuracy and maximality together implies correspondence between judgment and knowledge, $\mathsf{s}_p^{+/-} \cap \hat{S}^i = \hat{\mathsf{s}}_p^{+/-}$. The agent a_i's local state s_i being an element of $\mathsf{s}_p^{+/-}$ is a necessary and sufficient condition for $p/\neg p$ being true at that point in execution. We call these *positive* and *negative correspondence*, respectively. When the information availability condition holds together with both positive and negative correspondence, s_i must be either in s_p^+ or s_p^- and the agent has not only correct but complete information on p. We call this a *complete judgment.*

21.1.3 Action

In their analysis of actions, Israel, Perry and Tutiya (Israel et al. 1991, Israel et al. 1993) emphasized the distinction between movements and accomplishments. The latter is circumstantial, whereas the former is not. Their example was Brutus's killing of Caesar. Brutus's certain movement of his hand would bring about the death of Caesar, if relevant circumstantial conditions obtain, which would include, among other things, Brutus's holding a knife in his right hand, and Caesar's being within his arm's reach. A movement can be considered as an execution of a certain program by an agent which is independent from particular circumstances, whereas an accomplishment is an effect of that execution under a certain circumstance. We follow this distinction, and define acts and actions independently.

Acts are programs executed by agents[1]. Although programs were previously specified in terms of a protocol Π, we use an equivalent and more program-like definition of acts here.

Definition 3 [Acts] An *act* of an agent a_i is either an element of a set of basic acts $BACT_i$ for a_i, or a complex act constructed inductively from other acts α, β and a judgment φ of the form $s_i \in \mathsf{s}_p^{+/-}$ by the application of one of the operations below.

(1)	Sequences	$\alpha;\beta$
(2)	Conditionals	$\texttt{if}\ \varphi\ \texttt{then}\ \alpha\ \texttt{else}\ \beta$
(3)	Loops	$\texttt{repeat}\ \alpha\ \texttt{until}\ \varphi$

We indicate the set of all acts for a_i by ACT_i. The component $\texttt{CMP}(\alpha)$ of an act α is the set of all basic acts in α. Introduction of program-like structure into acts assumes a natural division of an agent's internal state into two components, a memory state component, which reflects the state of the environment, and a program control state component, which deter-

[1]Since our distributed-system based analysis provides us with only total events, we choose to make the act/action distinction correspond to the type/token distinction here, even though, strictly speaking, an act itself should be a token event in which a program is executed, and the program is a description of the corresponding act type.

mines what basic acts to execute next. We further assume here that the internal computation of agents including judgments are much faster than the execution of basic acts which brings about effects in the environment.

In contrast to the non-circumstantial nature of acts, actions include circumstantial effects of executions of acts.

Definition 4 [Actions] An *action* A is a binary relation R_{A} on the set of global states S.

$$R_{\mathrm{A}} \subseteq S \times S$$

We define the *restriction* of R_{A} on initial states with $\Delta \subseteq S$ to be

$$\Delta \rceil R_{\mathrm{A}} = R_{\mathrm{A}} \cap (\Delta \times S).$$

This definition of actions takes them as a set of transitions between global states. We define $R_{\mathrm{A}}^{\leftarrow}$ and $R_{\mathrm{A}}^{\rightarrow}$ to be the functions derived from R_{A}, which, given a final or an initial state, provide us with the set of initial or the set of final states, respectively, related with them by R_{A}.

$$\begin{aligned} R_{\mathrm{A}}^{\leftarrow}(s') &= \{s : R_{\mathrm{A}}(s, s')\} \\ R_{\mathrm{A}}^{\rightarrow}(s) &= \{s' : R_{\mathrm{A}}(s, s')\} \end{aligned}$$

We further define $S_{\mathrm{A}}^{\mathsf{PRE}}$ and $S_{\mathrm{A}}^{\mathsf{POST}}$ to be the set of initial states and the set of final states of transitions in R_{A}, respectively.

$$\begin{aligned} S_{\mathrm{A}}^{\mathsf{PRE}} &= \{s : \exists s'.R_{\mathrm{A}}(s, s')\}, \\ S_{\mathrm{A}}^{\mathsf{POST}} &= \{s' : \exists s.R_{\mathrm{A}}(s, s')\} \end{aligned}$$

If a proposition p holds in all of the states in $S_{\mathrm{A}}^{\mathsf{POST}}$, we call the action A an action of accomplishing p starting in $S_{\mathrm{A}}^{\mathsf{PRE}}$.

Actions may bring about changes not only in environmental states but also in agents' local states. If an action brings about in an agent an informational local state, we call the action a *test action.* When resulting states of an action S^{POST} satisfies the complete judgment condition, we call the action a *complete test.* Similarly, we call an action a *positive/negative test*, when S^{POST} satisfies positive/negative correspondence.

21.2 Implementation of Actions by Acts

When an act is executed in a certain environment, a certain result follows. Informally speaking, if a result of the execution of an act is an accomplishment of a resulting state of an action, then we can think of the act as implementing the action in that environment. To state it more formally, we first need to define the notion of behaviors of act executions.

21.2.1 Behaviors

We focus on the effects of acts of a particular agent a_i. When we fix a protocol for the environment (and for the other agents), an act α of a_i determines a set of (possibly infinite) sequences of global states, each element of which corresponds to a sequence of global states induced by possible executions of α by a_i. We call this set of global state sequences the behavior of α by a_i under that environment.

Definition 5 Let an *act mapping* ρ_{act} be a function from a global state to a set of global states that corresponds to a set of global state transitions induced by the execution of a basic act *act*.

$$\rho_{act} : S \to 2^S$$

Definition 6 We say a proposition p is *persistent* under a basic act *act* iff execution of *act* never falsifies p.

$$\forall s \in S. s \in S_p \supset (\forall s' \in \rho_{act}(s). s' \in S_p).$$

Let $\Gamma = (S \times ACT_i) \cup S$ be a set of configurations. A configuration $\gamma \in \Gamma$ represents the global state and the execution state of a_i's act. $\gamma \in S$ means the act execution has been terminated. We say a basic act *act* is the first element of α iff either α equals to *act*, or α equals to $act; \beta$ for some β, or α is of the form $\beta_1; \beta_2$ and *act* is the first element of β_1. We write γ^S for the global state of the configuration γ, and γ^N for the first element of the execution state of $\gamma \in S \times ACT_i$. We say γ enables a basic act *act* in states A and write $enable(\gamma, act, A)$, when $\gamma^N = act$ and $\gamma^S \in A$.

Definition 7 We define the *operational transition relation* $[\![\alpha]\!]$ over Γ for an act α by induction on the structure of α.

(1) BASIC ACTS

$$[\![\alpha]\!] \quad = \quad \{\langle\langle s, \alpha\rangle, s'\rangle : s' \in \rho_\alpha(s)\}$$

if α is a basic act.

(2) SEQUENCES

$$\begin{aligned}[\![\alpha;\beta]\!] \quad = \quad & \{\langle\langle s, \alpha'; \beta\rangle, \langle s', \alpha''; \beta\rangle\rangle : \langle\langle s, \alpha'\rangle, \langle s', \alpha''\rangle\rangle \in [\![\alpha]\!]\} \\ & \cup\{\langle\langle s, \alpha'; \beta\rangle, \langle s', \beta\rangle\rangle : \langle\langle s, \alpha'\rangle, s'\rangle \in [\![\alpha]\!]\} \cup [\![\beta]\!]\end{aligned}$$

(3) CONDITIONALS

$$[\![\texttt{if } \varphi \texttt{ then } \alpha \texttt{ else } \beta]\!] =$$

$$\{\langle\langle s, \texttt{if } \varphi \texttt{ then } \alpha \texttt{ else } \beta\rangle, \langle s, \alpha\rangle\rangle : s \in S_{\lambda s\, \varphi}\} \cup [\![\alpha]\!]$$
$$\{\langle\langle s, \texttt{if } \varphi \texttt{ then } \alpha \texttt{ else } \beta\rangle, \langle s, \beta\rangle\rangle : s \notin S_{\lambda s\, \varphi}\} \cup [\![\beta]\!]$$

(4) LOOPS

$$[\![\texttt{repeat } \alpha \texttt{ until } \varphi]\!] =$$

$$
\begin{aligned}
&\{\langle\langle s, \texttt{repeat } \alpha \texttt{ until } \varphi\rangle, \langle s, \alpha; \texttt{repeat } \alpha \texttt{ until } \varphi\rangle\rangle : s \notin S_{\lambda s.\varphi}\} \\
&\cup\{\langle\langle s, \alpha'; \texttt{repeat } \alpha \texttt{ until } \varphi\rangle, \langle s', \alpha''; \texttt{repeat } \alpha \texttt{ until } \varphi\rangle\rangle : \\
&\qquad \langle\langle s, \alpha'\rangle, \langle s', \alpha''\rangle\rangle \in [\![\alpha]\!]\} \\
&\cup\{\langle\langle s, \alpha'; \texttt{repeat } \alpha \texttt{ until } \varphi\rangle, \langle s', \texttt{repeat } \alpha \texttt{ until } \varphi\rangle\rangle : \\
&\qquad \langle\langle s, \alpha'\rangle, s'\rangle \in [\![\alpha]\!]\} \\
&\cup\{\langle\langle s, \texttt{repeat } \alpha \texttt{ until } \varphi\rangle, s\rangle : s \in S_{\lambda s.\varphi}\}
\end{aligned}
$$

Let $\boldsymbol{N}$ be a set of positive integers. A sequence σ of elements of a set E of length $n \in \boldsymbol{N}$ is a function from $I = \{i : 1 \leq i \leq n\}$ to E. We write $\sigma = [\sigma_1, \ldots, \sigma_n]$. σ is an infinite sequence, when $I = \boldsymbol{N}$. $seq^n\ E$ is a set of sequences of elements of E of length n. $seq^*\ E = \bigcup_{n \in \boldsymbol{N}} seq^n\ E$ is a set of finite sequences of elements of E. $seq^\omega\ E$ is a set of infinite sequences of elements of E, and $seq\ E = seq^*\ E \cup seq^\omega\ E$.

We would equate the behavior of an act with the set of sequences spanned by the transition relation $[\![\alpha]\!]$. But that set includes sequences in which a basic act keeps failing indefinitely without a success. We think those sequences are intuitively implausible to be regarded as among possible execution sequences of α, since we would assume that if it is possible for any basic act to succeed, then keep trying it would eventually turn out success and bring about the intended outcome. We will posit an additional condition on the sequences which excludes those infinitely failing sequences. The condition corresponds to the notion of fair scheduling of parallel processes (Manna and Pnueli 1992), and can be stated as follows: it is not the case that an act is performed infinitely many times but is succeeded only finitely many times.

Definition 8 A sequence $\sigma \in seq\ \Gamma$ is *fair* with respect to a basic act *act*, iff for any $A \subseteq S$ and $\Delta = \bigcup_{s \in A}\{\Delta_s : \phi \subset \Delta_s \subseteq \rho_{act}(s)\}$, the following does not hold.

$$\forall j \geq 0.\exists k \geq j.enable(\sigma_k, act, A) \ \wedge \ \exists j \geq 0.\forall k \geq j.{\sigma_k}^S \notin \Delta$$

We write $fair(\sigma, act)$ when σ is fair with respect to *act*, and define $fair(\sigma, \alpha) \equiv \forall act \in \texttt{CMP}(\alpha).fair(\sigma, act)$.

Definition 9 [Behaviors] The *finite complete behavior* of length $n \in \boldsymbol{N}$ for an act α starting in a state $s \in S$ is

$$
\begin{aligned}
\Sigma^n[\![\alpha]\!](s) \quad = \quad \{\sigma \in seq^n\ \Gamma : \ & \sigma_1 = \langle s, \alpha\rangle \\
& \wedge\ \forall i \in \{k : 1 < k \leq n\}.\langle\sigma_{i-1}, \sigma_i\rangle \in [\![\alpha]\!] \\
& \wedge\ \forall \gamma \in \Gamma.\langle\sigma_n, \gamma\rangle \notin [\![\alpha]\!]\}
\end{aligned}
$$

The *finite behavior* for α starting in s is

$$\Sigma^*[\![\alpha]\!](s) \quad = \quad \bigcup\{\Sigma^n[\![\alpha]\!](s) : n \in \boldsymbol{N}\}$$

The *infinite behavior* for α starting in s is

$$\Sigma^\omega[\![\alpha]\!](s) \quad = \quad \{\sigma \in seq^\omega\,\Gamma : \ \sigma_0 = \langle s, \alpha\rangle \wedge \forall i \in \boldsymbol{N}.\langle \sigma_i, \sigma_{i+1}\rangle \in [\![\alpha]\!] \wedge fair(\sigma, \alpha)\}.$$

The *behavior* of α starting in s is

$$\Sigma[\![\alpha]\!](s) \quad = \quad \Sigma^*[\![\alpha]\!](s) \cup \Sigma^\omega[\![\alpha]\!](s).$$

We can extend the notion of act mapping to general acts.

$$\rho_\alpha(s) = \begin{cases} \phi & \text{if } \Sigma^\omega[\![\alpha]\!](s) \neq \phi \\ \bigcup_{n\in\boldsymbol{N}}\{\sigma_n : \sigma \in \Sigma^n[\![\alpha]\!](s)\} & \text{otherwise} \end{cases}$$

$\rho_\alpha(s) \neq \phi$ ensures that all executions of α from s will terminate.

21.2.2 Implementation

We can now state the notions of implementation precisely.

Definition 10 [Implementation] An act α *implements* an action A iff

(1) every sequence in the behavior of α starting from states in initial states of the action A terminates
$\forall s \in S_{\mathrm{A}}^{\mathrm{PRE}}\ \rho_\alpha(s) \neq \phi$,

(2) and, each sequence ends in a state within the range of results of the action A
$\forall s \in S_{\mathrm{A}}^{\mathrm{PRE}}\ \rho_\alpha(s) \subseteq R_{\mathrm{A}}^{\rightarrow}(s)$.

Similarly, an act α *partially implements* an action A iff

(1) every sequence in the behavior of α starting from states in initial states of the action A terminates
$\forall s \in S_{\mathrm{A}}^{\mathrm{PRE}}\ \rho_\alpha(s) \neq \phi$,

(2) and, for each initial state there is at least one sequence that ends in a state within the range of results of the action A
$\forall s \in S_{\mathrm{A}}^{\mathrm{PRE}}\ \rho_\alpha(s) \cap R_{\mathrm{A}}^{\rightarrow}(s) \neq \phi$.

When an act α implements an action of accomplishing p, execution of α is guaranteed to terminate and bring about p. When α partially implements an action of accomplishing p, execution of α is also guaranteed to terminate, and it is not the case that execution of α never brings about p.

Definition 11 [Prefix-implementation] An act α *prefix-implements* an action A iff every sequence in the behavior of α starting from states in initial states of the action A eventually reaches a state within the range of the action A.

$$\forall s \in S_{\mathrm{A}}^{\mathrm{PRE}}.\forall \sigma \in \Sigma[\![\alpha]\!](s).\exists n \in \boldsymbol{N}.\ (\sigma_n = \langle s', \beta\rangle \vee \sigma_n = s') \wedge s' \in R_{\mathrm{A}}^{\rightarrow}(s).$$

Similarly, an act α *partially prefix-implements* an action A iff for each initial state of the action A, there is at least one sequence in the behavior of α that eventually reaches a state within the range of the action A.

$$\forall s \in S_{\mathrm{A}}^{\mathrm{PRE}}.\exists \sigma \in \Sigma[\![\alpha]\!](s).\exists n \in \boldsymbol{N}.$$
$$(\sigma_n = \langle s', \beta\rangle \vee \sigma_n = s') \wedge s' \in R_{\mathrm{A}}^{\rightarrow}(s).$$

When an act α prefix-implements an action of accomplishing p, execution of α may not terminate, but it is guaranteed to bring about p. When α partially prefix-implements an action of accomplishing p, execution of α, again, may not terminate, but it is not the case that execution of α never brings about p. It is clear from the definitions that implementation implies both partial and prefix-implementation, and both partial and prefix-implementation imply partial prefix-implementation.

The following property is also obvious from the definitions.

Proposition 1 If an act α implements an action R_{A}, then α implements its restriction on initial states $\Delta\rceil R_{\mathrm{A}}$ for any $\Delta \subseteq S$. Similar conditions hold for other types of implementation relations.

21.2.3 Implementation Conditions for Conditionals and Loops

Implementation conditions for composite acts are stated above in terms of their act mappings and behaviors they generate. Their definition is precise but relatively useless, since the definition doesn't take advantage of the power of test actions, which is the whole point of executing conditional and loop acts. More interesting conditions can be stated as relationship between implementation of composite acts and that of component acts.

CONDITIONALS

Proposition 2 If the agent has complete judgment for a proposition r at every state $s \in S_{\mathrm{A}}^{\mathrm{PRE}}$, α implements an action $S_r\rceil R_{\mathrm{A}}$ and β implements an action $S_{\neg r}\rceil R_{\mathrm{A}}$, then the conditional act `if` $s_i \in \mathbf{s}_r^+$ `then` α `else` β implements the action R_{A}. Similar conditions hold for other types of implementation relations.

Note that if the agent has only positive and not complete judgment for a proposition r in some state $s \in S_{\mathrm{A}}^{\mathrm{PRE}}$, even when α implements an action $S_r\rceil R_{\mathrm{A}}$ and β implements an action $S_{\neg r}\rceil R_{\mathrm{A}}$, the conditional act `if` $s_i \in \mathbf{s}_r^+$ `then` α `else` β may not even partially implement R_{A}, unless β itself implements $S_r\rceil R_{\mathrm{A}}$.

LOOPS

We can think of two different conditions corresponding to different types of loops. The first type is an accumulation of successful actions that results in an accomplishment of a desired state. To keep moving forward until you reach the destination would be an example of this type of the loop.

Proposition 3 If there is a sequence of sets of global states U_i $(i \geq 0)$ where proposition p holds in U_0 and the agent has complete judgment for p in every state in U_i, and α implements R_{A_i} where $S^{\mathrm{POST}}_{\mathrm{A}_{i+i}} \subseteq U_{i+1} \subseteq S^{\mathrm{POST}}_{\mathrm{A}_i}$ for all $i \geq 0$, then the loop act **repeat** α **until** $(s_i \in \mathbf{s}^+_p)$ implements an action R_{A} of accomplishing p, where $S^{\mathrm{PRE}}_{\mathrm{A}} = \bigcup_{i\geq 0}.U_i$, $S^{\mathrm{POST}}_{\mathrm{A}} = U_0$, and

$$R_{\mathrm{A}}(s,s') \Leftrightarrow (s = s') \vee (\exists k \geq 0. s_k = s \wedge s_0 = s' \wedge \bigwedge\nolimits_{j=o}^{k}.R_{\mathrm{A}_j}(s_{j+1}, s_j)).$$

The second type is of the loops in which possibly erroneous actions are iteratively attempted until the execution eventually ends in success. Search, both individual and joint, would be an example of this type of loops.

Proposition 4 If an act α partially implements an action R_{A} of accomplishing p where $S^{\mathrm{PRE}}_{\mathrm{A}} = U$ and $S^{\mathrm{POST}}_{\mathrm{A}} \subseteq U \cup S_p$, and it also implements a complete test on p in states in $S^{\mathrm{POST}}_{\mathrm{A}}$, and p is persistent under each basic act in $\mathtt{CMP}(\alpha)$, then the loop act **repeat** α **until** $(s_i \in \mathbf{s}^+_p)$ implements an action of accomplishing p and that of accomplishing $K^U_{a_i}p$ starting in U.

Proof. Since α partially implements R_{A} of accomplishing p, for any $s \in S^{\mathrm{PRE}}_{\mathrm{A}}$, $(\rho_\alpha(s) \cap S_p) \neq \phi$. So, if we take $\Theta_p = \{act \in \mathtt{CMP}(\alpha) : (\bigcup_{s\in S} \rho_{act}(s) \cap S_p) \neg \phi\}$, then $\Theta_p \neq \phi$. Let $A^{act} = \{s : \rho_{act}(s) \cap S_p \neq \phi\}$ and $\Delta^{act} = \bigcup_{s\in A^{act}}\{\rho_{act}(s) \cap S_p\}$ for each $act \in \Theta_p$. Then, for any $\eta \in \Sigma[\![\alpha]\!](s)$, there exists an $act \in \Theta_p$ and $k \geq 0$ such that $enable(\eta_k, act, A^{act})$. If we take an arbitrary execution sequence σ in the behavior $\Sigma[\![$**repeat** α **until** $(s_i \in \mathbf{s}^+_p)]\!](s)$ starting from a state $s \in S^{\mathrm{PRE}}_{\mathrm{A}}$, then either $enable(\sigma_k, act, A^{act})$ holds finitely many times for every $act \in \Theta_{U,p}$ or it holds infinitely many times for a certain act. In the first case, σ itself has to be of finite length, since α partially implements accomplishing of p, hence from the completeness of the judgment, $\exists k \geq 0.(\sigma_k \in S_p)$. In the second case, the fairness requirement on behaviors guarantees that there exists $k \geq 0$ such that ${\sigma_k}^S \in \Delta^{act}$, that is, ${\sigma_k}^S \in S_p$. From persistence of p and judgment completeness, it follows that for some $k' \geq k$, σ terminates at k' and $\sigma_{k'} \in S_p$. In both cases, if we let s to be the terminal state of σ, then the judgment $s_i \in \mathbf{s}^+_p$ holds in s. Since the judgment is complete in U, $K^U_{a_i}p$ holds in s. □

Proposition 5 If an act α partially implements an action R_{A} of accomplishing p where $S^{\mathrm{PRE}}_{\mathrm{A}} = U$ and $S^{\mathrm{POST}}_{\mathrm{A}} = U \cup S_p$, and p is persistent under each basic act in $\mathtt{CMP}(\alpha)$, but α only implements a positive and not a complete test on p in some states in U, then the loop act **repeat** α **until** $(s_i \in \mathbf{s}^+_p)$ prefix-implements the action of accomplishing p, but does not even partially prefix-implement the action of accomplishing $K^U_{a_i}p$ starting in U.

Proof. Similar to the proof of Proposition 4. □

21.3 Examples: INFORM as a Test

Let us examine several examples to see how our distributed-system model is applied to analyzing relationships between information and action in actual act executions. We think of several scenarios in each of which an agent performs a search either alone or jointly with other agents.

Let us first suppose that an agent a_1 is searching her lost purse alone by executing the act α below.

$$\alpha \stackrel{\mathrm{def}}{=} \texttt{repeat SEARCH until } (s_i \in \mathsf{s}^+_{\texttt{FOUND}_i})$$

SEARCH is a basic act for searching, and $\texttt{FOUND}_i$ is the proposition that a_i has found the purse. In this case $i = 1$. We can make following natural assumptions about executions of basic acts.

(A1) Execution of a single step SEARCH by a_i may or may not accomplish the discovery of the purse. So, SEARCH partially implements the action of accomplishing $\texttt{FOUND}_i$ starting in S.

$$\forall s \in S.(\rho_{\texttt{SEARCH}}(s) \cap S_{\texttt{FOUND}_i}) \neq \phi$$

(A2) a_i always knows whether she found the purse, that is, a_i always has a complete judgment on $\texttt{FOUND}_i$. Hence SEARCH is a complete test on $\texttt{FOUND}_i$.

$$\forall s \in S.(s_i \in \mathsf{s}^+_{\texttt{FOUND}_i}) \equiv (s \in S_{\texttt{FOUND}_i})$$

(A3) Execution of SEARCH does not lose the already found purse. So, $\texttt{FOUND}_i$ is persistent under SEARCH.

With these characterizations we can show from Proposition 4 that α implements both an action of accomplishing $\texttt{FOUND}_1$ and that of accomplishing $K^S_{a_1}\texttt{FOUND}_1$ starting in S. By executing α, a_1 will eventually discover the purse and she can know the discovery.

Now slightly modify the example, and assume a_1 is searching something not immediately known when it is found, a purse of her friend, on which she may need to perform some test action, like asking her friend, to identify it. The execution of α can be characterized similar to the case above, except that a_1 has only a positive correspondence and not a complete judgment on $\texttt{FOUND}_1$ for an arbitrary state in S. Under this characterization, we can show that α prefix-implements an action of accomplishing $\texttt{FOUND}_1$ starting in S, but does not even partially prefix-implements an action of accomplishing $K^S_{a_1}\texttt{FOUND}_1$. a_1 will eventually find the purse, but she may not notice the discovery.

Let us next suppose that a_1 executes α' below.

$$\alpha' \stackrel{\mathrm{def}}{=} \texttt{repeat SEARCH; IDENTIFY until } (s_1 \in \mathsf{s}^+_{\texttt{FOUND}_1})$$

We would assume the following on IDENTIFY.

(A2') Execution of IDENTIFY always gives a_1 the information whether she found the purse, that is, IDENTIFY implements a complete test on $\mathtt{FOUND}_1$.

$$\forall s \in \{\rho_{\mathtt{IDENTIFY}}(s) : s \in S\}.(s_1 \in \mathsf{s}^+_{\mathtt{FOUND}_1}) \equiv (s \in S_{\mathtt{FOUND}_1})$$

(A3') Execution of IDENTIFY does not lose the already found purse, that is, $\mathtt{FOUND}_1$ is persistent under IDENTIFY.

If we take $U = \{s : (s_1 \in \mathsf{s}^+_{\mathtt{FOUND}_1}) \equiv (s \in S_{\mathtt{FOUND}_1})\}$, it is easy to show that the act SEARCH; IDENTIFY both partially implements an action of accomplishing $\mathtt{FOUND}_1$ starting in U, and implements a complete test on $\mathtt{FOUND}_1$. Since (A3)(A3') states that $\mathtt{FOUND}_1$ is persistent under basic acts in α', Proposition 4 shows that α' implements both an action of accomplishing $\mathtt{FOUND}_1$ and that of accomplishing $K^U_{a_1}\mathtt{FOUND}_1$ starting in U. This example shows that by making a modification to her own act to execute, namely, the insertion of an act which implements a relevant test action, the agent introduces a restriction on available states, S to U, thereby making it possible to accomplish her goal with a loop act.

Next, we examine joint search cases. Imagine two agents a_1 and a_2 are jointly searching a lost purse. In a multi-agent case, other agents' program executions become effectively a part of the environment for each agent. We focus here on the agent a_1 and regard a_2 as part of the environment.

Let's first assume that both agents are executing essentially the same program α above synchronously. We need to distinguish the proposition $\mathtt{FOUND}_i$ that the purse is found by the agent a_i, from the proposition $\mathtt{FOUND} = \mathtt{FOUND}_i \vee \mathtt{FOUND}_j$ that the purse is found by either of the two agents. The goal of joint search should be to bring about the latter, whereas each agent a_i can attain knowledge only about the former. From the nature of SEARCH act, it would be natural to assume that SEARCH partially implements an action of accomplishing FOUND[2]. However, the judgment $s_i \in \mathsf{s}^+_{\mathtt{FOUND}_i}$ is only a positive correspondence and not a complete judgment on FOUND. From Proposition 5, together with the persistence of FOUND under SEARCH, it can be shown that the act α prefix-implements an action of accomplishing FOUND starting in S, but it does not even partially prefix-implement an action of accomplishing $K^S_{a_i}\mathtt{FOUND}$. This corresponds to a scenario where an agent keeps searching without knowing that the other agent already found the purse.

Now, suppose each agent synchronously executes β below.

[2]SEARCH may not partially implement an action of accomplishing $\mathtt{FOUND}_i$, when there is only one purse to be found.

$\beta \stackrel{\text{def}}{=} \texttt{repeat } \beta_0 \texttt{ until } (s_i \in \mathrm{s}^+_{\texttt{FOUND}})$

where $\beta_0 \stackrel{\text{def}}{=} \texttt{SEARCH; if } (s_i \in \mathrm{s}^+_{\texttt{FOUND}_i}) \texttt{ then INFORM else NO-OP}$

Agent a_i has her own judgment on FOUND, which is different from the judgment on $\texttt{FOUND}_i$.

$$\mathrm{s}^+_{\texttt{FOUND}} \stackrel{\text{def}}{=} \mathrm{s}^+_{\texttt{FOUND}_i} \cup \mathrm{s}^+_{\texttt{INFORMED}_{ij}}$$

$\texttt{INFORMED}_{ij}$ stands for the proposition that a_i is informed from a_j of her search result. Note that $\texttt{INFORMED}_{ij}$ need not imply $\texttt{FOUND}_j$ when a_j may not be truthful. We further assume that the following intuitively plausible constraints hold on the executions of the basic acts by a_i.

(B1) Execution of INFORM by a_i is the infallible and also the only way to bring about that a_j is informed.

$$\{\rho_{\texttt{INFORM}}(s) : s \in S\} = \{s : s_j \in \mathrm{s}^+_{\texttt{INFORMED}_{ji}}\}$$

(B2) a_i always knows whether she is informed, that is, a_i always has a complete judgment on $\texttt{INFORMED}_{ij}$.

$$\{s : s_i \in \mathrm{s}^+_{\texttt{INFORMED}_{ij}}\} = \{s : s \in S_{\texttt{INFORMED}_{ij}}\}$$

(B3) $\texttt{FOUND}_i, \texttt{FOUND}_j, \texttt{INFORMED}_{ij}, \texttt{INFORMED}_{ji}$ are all persistent under execution by a_i of any of SEARCH, INFORM and NO-OP, and only SEARCH executed by a_i brings about $\texttt{FOUND}_i$.

If we take $U = \{s : s_i \in \mathrm{s}^+_{\texttt{FOUND}_i} \equiv s_j \in \mathrm{s}^+_{\texttt{INFORMED}_{ji}} \text{ for all } i\}$, we can show that the act β_0 both partially implements an action of accomplishing FOUND starting in U, and implements a complete test on FOUND. Together with persistence of FOUND under basic acts, we can rely again on Proposition 4 to show that β implements both an action of accomplishing FOUND and that of accomplishing $K^U_{a_i}\texttt{FOUND}$ starting in U.

This example demonstrates that in a multi-agent setting a modification in one's own act can bring about a change in the behavior of environment for other agents, which, in turn, introduces a restriction on available states, S to U, thereby making it possible to accomplish one's goal with a loop act. In our example, INFORM act effectively works as a test action for other agents. We can view this type of arrangements as implementing a test action by not choosing one's own action but by changing the behavior of one's own environment. By agreeing among several agents on what acts to execute, they can effectively modify their environment, and thereby establishing a different type of methods of accomplishing their goals.

21.4 Conclusions

We proposed a model for situated agent actions based on the framework of distributed systems. The model makes a clear distinction between environment-dependent notions, knowledge and actions, from environment-independent notions, judgments and acts. These distinctions make it possible to obtain descriptions either from theorists/designers' perspectives or from agents' perspectives. Secondly, the model captures the effects of acts executed in an environment in terms of implementation relationships between acts and actions. Thirdly, it leaves room for judgment uncertainty, which enables us to talk of the difference between acts triggered by certain information and those triggered by uncertain information, which eventually gives us a distinction between implementation and prefix-implementation. Finally, as we demonstrated in joint search examples, the model provides a natural method of analyzing multi-agent actions and gives a unified conception to information oriented actions.

References

Cohen, Philip R., and Hector J. Levesque. 1990. Persistence, intention and commitment. In *Intentions in communication*, ed. Philip R. Cohen, Jerry Morgan, and Martha E. Pollack. Chap. 3, 33–70. MIT Press.

Halpern, Joseph Y., and Yoram Moses. 1990. Knowledge and common-knowledge in a distributed environment. *Journal of the ACM* 37(3):549–587.

Israel, David, John Perry, and Syun Tutiya. 1991. Actions and movements. In *Proceedings of IJCAI-'91*. Morgan Kaufmann.

Israel, David, John Perry, and Syun Tutiya. 1993. Executions, Motivations and Accomplishments. *Philosophical Review* 102(4):515–540.

Levesque, Hector J., Philip R. Cohen, and J. Nunes. 1990. On acting together. In *AAAI-90*, 94–99. Morgan Kaufmann.

Manna, Zohar, and Amir Pnueli. 1992. *The temporal logic of reactive and concurrent systems: specification*. Springer-Verlag.

Moore, Robert C. 1985. A formal theory of knowledge and action. In *Formal theories of the commonsense world*, ed. Jerry R. Hobbs and Robert C. Moore. Chap. 9, 319–358. Ablex Publishing Corporation.

Perrault, C. Raymond. 1990. An application of default logic to speech act theory. In *Intention in Communication*, ed. Philip R. Cohen, Jerry L. Morgan, and Martha E. Pollack. 105–133. MIT Press.

Rao, Anand S., and Michael P. Georgeff. 1991. Modeling rational agents within a BDI-architecture. In *Proceedings of KR-'91*, 473–484. Morgan Kaufmann.

22

Information, Representation and the Possibility of Error

ROBERT C. KOONS

Introduction

Attempts to explicate the phenomenon of representation naturalistically, say, in terms of causal connection, often founder on the problem of explaining the possibility of error. Suppose, for example, that we attempt to explain representation along these lines: fact σ *represents* fact τ iff τ is a causally necessary condition of σ. Such an account, of course, leaves no room for the possibility of misrepresentation, if 'necessary condition' is interpreted strictly. Whenever an actual σ represents a state of affairs τ, the state τ will be actual and σ will not be in any way a misrepresentation of the world.

A common strategy for solving this problem is to distinguish between two types of situation: type 1 and type 2 (see Fodor 1990, p. 60). We can then identify the content of a representation with that fact which is causally necessitated by the form of the representation in type 1 situations. A representation in a type 2 situation can then misrepresent the world, since its content is determined with reference to a counterfactual situation: what would be causally necessitated by the form of the representation were it to be located in a situation of type 1 instead of type 2.

The distinction between type 1 and type 2 situations is typically made in terms of the historical antecedents of the representational form. For example, in Dretske's account of representation Dretske 1981, type 1 situations are those situations that occur during the training period in which the meanings of the representational forms are impressed upon the individual subject. Similarly, in Millikan's account Millikan 1984, type 1 situations

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

are those situations that actually occurred in the evolutionary history of the representational system in question.

The strategy of appealing to historical antecedents leads to a number of serious difficulties, as I will argue in section 1. Firstly, the historical strategy tends to attribute contents that are far too weak, since misrepresentations do in fact occur in type 1 situations, and these are misdescribed as veridical whenever the historical strategy is followed. Secondly, the historical strategy makes content far too sensitive to irrelevant accidents of history. Finally, this strategy would force us to make facts about the remote past relevant to the best theoretical account of the present.

In Section 2, I develop a naturalistic account of the nature of information and of the possibility of error that avoids the historical strategy, the type 1/type 2 distinction, and the concomitant difficulties. This account relies exclusively on probabilistic connections between states, interprets information in terms of probabilistic necessitation, and leaves room for the possibility of error by failing to make the erroneous inference (made, surprisingly enough, by both Dretske and Fodor) from an event's having probability *zero* (or infinitely close to *zero*) to that event's being absolutely impossible. I develop a conditional logic, building on work on counterfactuals by Stalnaker (1986), Lewis (1973), and Skyrms (1980), in which the conditional $\phi \Box\!\!\rightarrow \psi$ represents a conditional probability of ψ on ϕ that is infinitely close to *one*, i.e., some infinitesimal distance from *one*. In addition, I use this conditional logic to build a principled nonmonotonic consequence relation that is related to the theory of commonsense entailment constructed by Asher and Morreau (1991). These logics are then used to define an information linkage relation and a related notion of *natural error* or *misinformation*.

In Section 3, I construct a definition of the causal nexus and of causal priority, using the tools developed in Section 2. It is at this point that situation theory plays a crucial role. Natural functions are shown to correspond to certain nested causal laws. I argue that Millikan is right in thinking that Darwinism provides a scientific basis for teleology and does not, as is often assumed, banish teleology from nature.

Finally, in section 4, I sketch a new account of the nature of representation and perception. I take perceptual representation and motor-volitional representation to be special cases of information: a fact σ *represents*$_0$ a state of affairs τ in an organism of type μ just in case σ has the *natural function* in μ of carrying the information that τ is actual. Perceptual misrepresentation is, in turn, a special case of natural error or misinformation. Cognitive representations, like desires, intentions, doubts, and imaginings, derive their content from modality-specific functional connections with perceptual and motor representations.

I explain how causation is needed in specifying the semantics of naked-infinitive perception reports, and how my definition of causation accounts for the logical properties of these reports that were first observed by Barwise and Perry (1983).

22.1 The Historical (Retrospective) Strategy

The simplest information-based theory of representation would go something like this: a representation type σ represents the actuality of some state of affairs τ just in case it is causally impossible that σ be actual without τ's being actual. This means that either τ is a causally necessary condition for σ or σ is a causally sufficient condition for τ. Unfortunately, this simple theory attributes content to representation types that is far too weak, so weak that error is impossible. On this account, if σ is actual, and σ represents τ, then τ must also be actual.

One way to narrow the content of a representation-type and thereby to explain the possibility of error is to appeal, not only to the present causal properties of the representation-type, but also to facts about the actual history of the type (even its remote history). These facts may be facts about the previous history of the individual symbol user (as in Dretske's (1981) theory or about the history of the representational practice to which the type belongs (as in Millikan's (1984) account). A simple Dretske-like theory might take the following form: representation-type σ represents τ for subject A iff for every situation s belonging to the training period (during which A learned the meaning of σ), τ was causally necessary *in the circumstances* (in s) for σ. We could say that a state of affairs τ is causally necessary in the circumstances of s for σ iff there is a state of affairs υ such that both σ and υ are actual in s and τ is causally necessary for the joint occurrence of σ and υ. Misrepresentation is possible for any representation occurring outside of the training period, because an token of *sigma* occurring in a situation s outside the training period might represent τ even though τ is not necessary for σ in the circumstances of s.

This sort of account has the paradoxical result that the longer and more varied the trainig period, the weaker the content of the representation-type. If, for example, we extended the scope of the relevant history to include the whole history of the representational practice (in the case of natural representations, this would mean the entire evolutionary history of the species), the resulting content would be so weak as to render error virtually impossible.

An alternative but very simple informational account would stipulate that σ represents τ just in case the occurrence of σ increases the objective probability of τ. Let's say that σ *probabilifies* τ just in case the objective conditional probability of τ on σ is greater than that of τ on the dual of σ. We could say that σ represents τ just in case σ probabilifies τ. This account has a defect that is exactly opposite to the defect we encountered in the simple causal-necessitation model. Instead of making error impossible, this account makes error absolutely ubiquitous. Every representation represents innumerably many possible states of affairs, all but a vanishingly small proportion of which are nonexistent.

Millikan starts with this probabilizing model and solves the ubiquity of error problem by adopting a version of the historical strategy. On a simple Millikan-like account, we could stipulate that σ represents τ iff σ probabilifies τ, and the fact that σ probabilifies τ has in reality contributed causally to the perpetuation of some reproductive family to which σ belongs. The longer and more varied is the relevant evolutionary history, the narrower are the contents ascribed to the representation and the more frequent are the errors and misrepresentations. Indeed, as Millikan recognizes, there will be many representational forms that will be erroneous on nearly every occasion (Millikan 1984, p. 34). For example, suppose that some pattern of auditory stimulation increases the probability of the presence of a predator, and that this pattern has triggered a flight response in the past, contributing thereby to the perpetuation of the species. Then, on Millikan's account, the auditory pattern represents "Predator near!", even though on nearly all occasions, the pattern is caused by the wind's rustling of leaves. In fact, Millikan's account cannot provide a basis for ascribing probabilistic content. For example, we could not, on her account, distinguish between signals that mean "There's a slight chance of a predator near" from those that mean "More likely than not there's a predator near" or "Without a doubt a predator is near". All of these signals would simply represent "Predator near" without qualification.

There are a number of other difficulties that could be raised concerning the details of Dretske's theory or of Millikan's, but here I would like to concentrate on some problems that are endemic to the historical strategy itself. Firstly, reliance on the historical strategy causes deviant cases in the past to influence the content of representations. For instance, it is quite common for training periods to include some cases in which the representation is wrongly but plausibly applied. I could teach a child the true meaning of 'bird' by means of cleverly constructed mechanical models, even though every attribution of the term in the training period was false. Similarly, in the evolutionary history of any representational system, there will be events in which misrepresentations accidentally contributed to the survival of the system.

Secondly, the historical strategy makes content too sensitive to accidental features of history. For the sake of illustration, consider the following version of the Twin Earth thought-experiment. Suppose that on Twin Earth, both H_2O and XYZ occur in equal abundance, and in close proximity to one another: here an H_2O lake, there an XYZ river, and so on. Suppose further that, simply as a matter of pure coincidence, the inhabitants of Twin Earth have encountered only H_2O and have applied to it the term 'water'. Applying the historical strategy means interpreting this symbol as designating only H_2O, despite the fact that Twin Earthers are, in the future, just as likely to encounter XYZ as H_2O and are completely unable to discern any difference between the two.

Thirdly, the historical strategy makes facts about the remote past directly relevant to the ascription of content to present-day representations. Content ascription should enable us to understand and explain the behavior of rational agents; information about the remote past of such agents cannot be of any immediate significance for this task, unless we are to believe in something like action at a temporal distance.

22.2 A New Strategy

Information is somehow tied to objective probabilistic relevance. In explicating this tie, we seem to be faced with a dilemma. If we insist that whenever a fact σ carries the information τ, the objective conditional probability of τ on σ be *one*, then we make σ sufficient for τ, thereby eliminating any possibility of error. Alternatively, if we require only that the conditional probability of τ on σ be very high (though not necessarily equal to *one*), then we run afoul of a very important principle of information: what Dretske calls the Xerox Principle (Dretske 1981, pp. 57–8).

Dretske's Xerox Principle is simply the requirement that the carriage of information is transitive: if σ carries τ, and τ carries υ, then σ carries υ. Obviously, if we set some finite distance ϵ from *one* as the threshhold on conditional probability for the carriage of information, then this carriage will not be transitive.

The dilemma stands only if one assumes (as Dretske explicitly does – Dretske 1981, p. 245) that it is impossible that a state have probability *one* and fail to be actual. This assumption is false for standard interpretations of the probability calculus, in which events of measure zero are quite possible. I agree with Dretske that this assumption is a useful one. However, one can accept this assumption and still avoid the dilemma, by using a non-standard probability theory, one permitting hyperreal, i.e., infinitesimal, quantities. Let $\sigma \rightsquigarrow \tau$ represent the fact that σ carries the information that τ. Let $\sigma \,\Box\!\!\rightarrow \tau$ represent that fact that the objective conditional probability of τ on σ is infinitely close (some infinitesimal distance from) *one*, and let $\sigma \,\Diamond\!\!\rightarrow \tau$ represent the fact that the objective conditional probability of τ on σ is **not** infinitely close to *zero*. I can then define information in terms of extreme conditional probabilities:

$$\sigma \rightsquigarrow \tau \leftrightarrow [\sigma \,\Box\!\!\rightarrow \tau \,\&\, \tau \,\Diamond\!\!\rightarrow \sigma]$$

This definition clearly satisfies the Xerox Principle, and, in addition, it is quite possible for σ to carry the information that τ even though σ is actual and τ is not, i.e., it is quite possible for a state to carry misinformation or natural error.

This account has the counterintuitive result that misinformation or natural error can be expected to occur with only an infinitesimal frequency. Two things can be said in response. First, since information is ubiquitous, the fact that the limiting relative frequency of misinformation is infinitesimal does not entail that the absolute frequency of error is low. Moreover, when misinformation is detected, this fact is especially vivid and salient, while the background of accurate information is taken for granted and largely unnoticed. Second, the usefulness of my account does not depend on taking the requirement of an infinitesimal relative frequency of error **literally**. Presumably, misinformation is exceptional, occurring with a very low relative frequency. At some point, very low finite probabilities are treated, for all practical purposes, as though they were infinitesimal. There are fairly obvious computational advantages to working with qualitative differences, represented formally as infinite ratios, instead of working exclusively with quantitative differences. What I am offering is a formal model of how we reason commonsensically about information. If the account faithfully reproduces the crucial features of our commonsense practice, then the question of its literal truth is of little or no importance. In actual practice, we apply descriptions like 'misinformation' or 'error' to cases of which the descriptions are not literally true, as, for example, we apply descriptions like 'flat' to surfaces that are not literally flat but are close enough to flatness for practical purposes.

It is possible to develop a formal semantics for the informational conditionals $\Box\!\!\rightarrow$ and $\Diamond\!\!\rightarrow$, building on work in the Kripkean tradition by Stalnaker Stalnaker 1986 and Lewis Lewis 1973, and on work on higher order probabilities by Skyrms Skyrms 1980 and Gaifman Gaifman 1986. I should also mention early work on conditionals and extreme probabilities by Ernest Adams, , and more recent work that makes explicit use of infinitesimal probabilities by SpohnSpohn 1988. In this abstract, I will make mention of some of the more significant features of this logic. A model $\mathcal{M}$ for the conditional logic consists of a tuple $\langle \mathcal{W}, \mu, \mathcal{I} \rangle$, where $\mathcal{W}$ is the nonempty set of worlds, μ assigns to each world w a probability function μ_w that assigns hyperreal numbers from 0 to 1 to subsets of $\mathcal{W}$, and $\mathcal{I}$ is an interpretation function that assigns the values T and F to ordered pairs consisting of worlds and atomic formulas. The truth conditions for the conditionals are the following:

$$\mathcal{M}, w \models \phi \,\Box\!\!\rightarrow \psi \iff \text{either } \mu_w(\| \phi \|) = 0 \text{ or } \frac{\mu_w(\| \phi \,\&\, \psi \|)}{\mu_w(\| \phi \|)} \approx 1$$

$$\mathcal{M}, w \models \phi \,\Diamond\!\!\rightarrow \psi \iff \mu_w(\| \phi \|) \neq 0 \text{ and } \frac{\mu_w(\| \phi \,\&\, \psi \|)}{\mu_w(\| \phi \|)} \not\approx 0$$

A complete axiomatization for this logic is essentially that of Lewis's familiar *VC* logic for counterfactuals, minus the modus ponens and strong centering for the counterfactual, and plus two additional axiom schemata, which I call the Skyrms axioms. In this logic, which I shall call *LSK*, the conditional is not stronger than the material conditional, as is the case for *VC*. The Skyrms axioms incorporate certain constraints on higher order probabilities in the models. They are restricted forms of what is often called absorption.

$$(\phi \,\Box\!\!\rightarrow (\psi \,\Box\!\!\rightarrow \chi)) \leftrightarrow (\phi \,\&\, \psi) \,\Box\!\!\rightarrow \chi$$

$$(\phi \,\Box\!\!\rightarrow \neg(\psi \,\Box\!\!\rightarrow \chi)) \leftrightarrow \neg((\phi \,\&\, \psi) \,\Box\!\!\rightarrow \chi)$$

where ϕ is in each case a Boolean combination of $\top$ and $\Box\!\!\rightarrow$-formulas. The Skyrms axioms are the logical analogue of what Skyrms called "Miller's principle", which states that the first-order probability weights can be recovered from higher order probabilities through integration. As Skyrms explains, a Dutch book argument can be given for Miller's principles, indicating that they are conditions on coherent agency, as are the familiar axioms of the probability calculus. The Skyrms axioms have a number of interesting implications for conditional logic. The following theorems of the conditional logic exploit some of the power of the Skyrms axioms:

Theorem 1 $\quad LSK \models (\top \Box\!\!\rightarrow (\phi \Box\!\!\rightarrow \psi)) \leftrightarrow (\phi \Box\!\!\rightarrow \psi)$

$LSK \models \Box\phi \rightarrow \Box\Box\phi$

$LSK \models (\phi \,\&\, (\phi \Box\!\!\rightarrow \psi)) \Box\!\!\rightarrow \psi$

We can use this conditional logic to define a form of nonmonotonic consequence, a form closely related to the theory of commonsense entailment of Asher and Morreau (1991), and to recent work by Pearl (1988) and Lehmann and Magidor (1992). The set of probability values is a set of hyperreals between 0 and 1. I assume that this set is so structured that every set X of probability values is *quasi-well-founded*, by which I mean that X has a member b such that no member of X is *infinitely* smaller than b. (If this assumption is relaxed, the resulting nonmonotonic logic lacks the property of rational monotonicity, but is otherwise well-behaved.)

$\Gamma \approx\!\!\!\models \phi \iff$ for almost all probability functions μ, $\mu(\phi/\Gamma) \approx 1$

Thus, nonmonotonic inference consists simply in updating a family of a priori probability function by Bayesian conditionalization, drawing a conclusion whenever almost all of these conditionalized probability functions agree in assigning the conclusion a probability infinitely close to 1. The standard of nonmonotonic correctness can be supported by the usual appeals to dynamic coherency made on behalf of Bayesianism.

We can prove a representation theorem for this nonmonotonic logic.

Definition 1 $\langle \phi, \psi \rangle$ constitutes a bf 1st order anomaly in $\mathcal{M}$ iff $\mathcal{M} \models \phi \& \psi \& (\phi \Box\!\!\rightarrow \neg\psi)$.

Definition 2 $\langle \phi, \psi \rangle$ constitutes an $n + 1^{st}$ order anomaly in $\mathcal{M}$ iff $\langle \phi, \psi \rangle$ constitutes a 1st order anomaly and there exists an n^{th} order anomaly $\langle \sigma, \tau \rangle$ in $\mathcal{M}$ such that $\mathcal{M} \models (\phi \Box\!\!\rightarrow (\sigma \& \tau))$.

Definition 3 The *anomaly level* of a model $\mathcal{M}$ is the greatest ordinal α such that $\mathcal{M}$ has an α-order anomaly.

Theorem 2 $\Gamma \mid\!\approx \phi \iff$ *for every $\mathcal{M}$ an anomaly-level minimizing model of Γ, $\mathcal{M} \models \phi$.*

The resulting logic is a rational consequence relation (in Delgrande's sense).

A bonus of building the nonmonotonic logic upon a modal conditional logic is that the nonmonotonic principle of Specificity is a consequence of independently motivated semantical principles and does not have to be secured in an ad hoc fashion, e.g., by imposing extrinsic priorities upon the default rules. Specificity is the principle that defaults with more specific conditions take precedence over those with less specific conditions.

The addition of a nonmonotonic consequence relation and the resources of Barwise-Perry situation theory makes it possible to define a notion of *robust* or *knowledge-bearing* information. A situation-theoretic model structure adds a set $\mathcal{S}$ of situations, a superset of $\mathcal{W}$ and a subsituation relation $\sqsubseteq$, a partial ordering of $\mathcal{S}$. The set $\mathcal{W}$ consists of the $\sqsubseteq$-maximal members of $\mathcal{S}$. In the canonical model, a situation can be modelled as a union of a set of literals and a deductively closed set of $\Box\!\!\rightarrow$-formulas and negations of these, and the $\sqsubseteq$ relation can be modelled by the subset relation. A situation s *robustly* carries the information τ in w ($w \models s \mapsto \tau$) just in case every supersituation s' of s that is a subsituation of the world w has τ as one of its nonmonotonic consequences. A fact σ *robustly* carries the information τ in w iff there is a set Γ of modal formulass true in w such that $\Gamma \cup \{\sigma\} \mid\!\approx \tau$, and for every ϕ such that $\Gamma \cup \{\sigma\} \mid\!\approx \phi$, $w \models \phi$.

Like simple information carriage, robust or knowledge-bearing information carriage is transitive (respects the Xerox Principle).

22.3 Causation and Teleology

Not all information is representation. Information pervades the inanimate world, while only organisms and machines produce and use representations. Millikan was right in thinking teleological explanation to be crucial to the explication of representation, but she was wrong to think that teleological explanation is retrospective, ineluctably tied to the historical strategy. In this section, I develop a fully prospective account of teleology, an account which is, moreoever, fully compatible with a physicalistic naturalism.

To begin, I must introduce the notions of *causation* and *causal direction*. Here I build on some recent work by Judea Pearl (Pearl 1988, pp. 116-133, 396-408). Pearl argues that causation introduces a strict partial ordering of atomic propositions. An acceptable ordering, given a fixed probabilistic model, is one that ensures that the immediate ancestors of a proposition render that proposition probabilistically independent of all of its non-descendants. The simplest acceptable orderings for a model constitute the best candidates for causal priority. In order to apply these insights to the nonmonotonic conditional logic developed in the last section, let $\prec$ be a strict partial well-ordering of atomic formulas. Let us say that an atomic formula α is a *constituent* of a formula ϕ relative to model $\mathcal{M}$ if α occurs in a simplest normal disjunctive form whose interpretation in $\mathcal{M}$ is identical to the interpretation of ϕ. We can then extend define two partial well-orderings, $\prec^+$ and $\prec^-$, which relate complex formulas:

$$
\begin{aligned}
\phi \prec^+ \psi \quad &\Longleftrightarrow \quad \exists\alpha\exists\beta(\alpha \text{ is a constituent of } \phi \\
&\qquad\quad \& \, \beta \text{ is a constituent of } \psi \,\&\, \alpha \prec \beta) \\
\phi \prec^- \psi \quad &\Longleftrightarrow \quad \forall\alpha(\alpha \text{ is a constituent of } \phi \rightarrow \exists\beta(\beta \text{ is a constituent of } \psi \\
&\qquad\quad \& \, \alpha \prec \beta)) \,\&\, \neg\exists\alpha\exists\beta(\alpha \text{ is a constituent of } \phi \\
&\qquad\quad \& \, \beta \text{ is a constituent of } \psi \,\&\, \beta \prec \alpha)
\end{aligned}
$$

Given the ordering $\prec^-$, I can define a causal conditional $\Box\!\!\Rightarrow$:

$$
\begin{aligned}
\mathcal{M}, w \models \phi \Box\!\!\Rightarrow \psi \quad &\Longleftrightarrow \quad \exists\chi(\phi \prec^- \chi \,\&\, \mathcal{M}, w \models \phi \Box\!\!\rightarrow \chi \\
&\qquad\quad \& \, \| \chi \|_{\mathcal{M}} \subseteq \| \psi \|_{\mathcal{M}})
\end{aligned}
$$

We can now impose the qualitative analogue of the independence condition Pearl argues to be essential to the idea of causal priority. A partial ordering $\prec$ is *acceptable* for a model-world pair $\langle \mathcal{M}, w \rangle$ just in case for every atomic formula α, if ϕ is a formula containing complete information about the immediate $\prec$-predecessors of α, and ψ is a formula of such a kind that $\alpha \not\prec^+ \psi$, then:

$$\mathcal{M}, w \models (\phi \,\&\, \psi) \Box\!\!\rightarrow \alpha \Longleftrightarrow \mathcal{M}, w \models \phi \Box\!\!\rightarrow \alpha$$

If a candidate causal ordering is acceptable for a model-world pair, then a limited form of strengthening of the antecedent of $\Box\!\!\rightarrow$ conditionals is truth-preserving: if the antecedent contains complete information about the immediate predecessors of the consequent, then any non-descendant formula may be added to the antecedent without affecting the conditional's truth-value at that world.

Given a nonmonotonic logic and a family of acceptable orderings (one for each world), we can define causal connections of two kinds: a situation as cause of a fact, and a fact as cause of another fact. Let's say that a situation s is causally prior to a fact σ in world w iff some fact supported by s precedes σ (according to $\prec_w$), and no fact supported by s is a descendant of σ (according to $\prec_w$). I will stipulate that situation s causes σ in w ($w \models s \rhd \sigma$) just in case s is a subsituation of w, s is causally prior to σ in w, s does not support σ, and every situation s' such that $s \sqsubseteq s' \sqsubseteq w$ has σ as one of its nonmonotonic consequences. The point of the last clause is to ensure that s is prima facie sufficient to bring about σ, and there is no fact in w outside of s that would override or counteract that sufficiency. In other words, s has a defeasible sufficiency for σ that is not in fact defeated in w.

A fact σ is a cause of a fact τ just in case σ is a necessary part of some situation s that causes τ. This definition captures the central insight of Mackie's INUS definition of cause: an insufficient but necessary part of an unnecessary but sufficient condition (Mackie 1965).

Definition 4 $\mathcal{M}, w \models \sigma \rhd \tau \iff \exists s(\mathcal{M}, w \models s \rhd \tau \,\&\, \forall s' \sqsubseteq s(s' \not\models \sigma \rightarrow \mathcal{M}, w \not\models s' \rhd \tau))$

Although teleological and functional explanations have been much discussed since the time of Plato, there is relatively little available on the question of the logical form of teleological laws. I conjecture that a teleological law consists of a certain kind of claim linking three generic or parametric facts or types: σ has the function of making it the case that τ in species or natural kind υ. As has often been noticed, teleological explanation seems to reverse the normal causal order: τ is the final cause of σ in υ, even though σ is causally prior (in the ordinary sense) to τ. This has very often seemed paradoxical, but the air of paradox disappears once the logical form of the teleological claim is seen to involve the assertion of a higher order causal law. To say that τ is the final cause of σ in υ is to claim that there is a higher order causal law whose antecedent contains the fact that something is an instance of υ and the fact that there is a causal law linking σ to τ, and whose consequent is τ itself. More formally,

Definition 5 $\quad T(\sigma, \tau, \upsilon) \iff [\upsilon \,\&\, \forall y(\neg\sigma[y/x] \Box\!\!\Rightarrow \neg\tau[y/x])] \Box\!\!\Rightarrow \sigma)]$

Teleological explanation seems paradoxical because τ occurs in the antecedent of the causal law and σ occurs in the consequent even though σ is causally prior to τ. This is not in fact a semantic irregularity, because τ does not occur in its own right in the antecedent; instead it occurs as part of a causal conditional. Although σ is causally prior to τ, it is not causally prior to $\forall y(\neg\sigma \Box\!\!\Rightarrow \neg\tau)$.

Consider a concrete example of a teleological connection. Suppose we claim that the function of a robin's tail is aerial stability. Let σ be the property of having a tail, τ be the property of aerial stability, and let υ be the property of being a robin. The teleological claim consists of a claim that there is a higher order causal law according to which the joint fact of something's being a robin and its being the case that having a tail is a causally necessary condition of aerial stability is a cause of that thing's having a tail. But there is nothing mysterious about such a higher order law. Such a law is a corollary of a Darwinian theory of natural selection. Darwinism is best understood, not as the thesis that there are no final causes in nature, but as the hypothesis that all final causes in nature are ultimately explicable in terms of reproductive advantage. Assuming that aerial stability is an adaptive feature of robins and that having a tail is indeed causally necessary (in the case of robins) for aerial stability, then this causal connection between tails and aerial stability is part of the explanation for actual robins' having tails: had their ancestors not acquired tails, robins would not have successfully reproduced.

Isn't this merely a variation of Millikan's retrospective strategy? No. The explanation for why only reproductively well-adapted organisms occur in significant numbers does indeed make reference to the past: only well-adapted species can be expected to have significant longevity, and, at any point in time, one can expect the vast majority of organisms to be members of long-lived species. However, the definition of the functions of a species does not in itself make any reference to the history of that species (as in Millikan's theory) , but only to the relatively high objective probability here and now of the successful reproduction of that species.

22.4 Representation and Perception

The basic form of representation is functional information. A parametric state σ represents$_0$ a state τ in natural kind υ if and only if it is the function of σ in υ to carry the information τ. Sense perceptual states are representations$_0$, as are motor volitions. Perceptual knowledge occurs when a perceptual state carries the information it represents **robustly**.

Naked-infinitive perceptual reports, first studied by Barwise and Perry (Barwise and Perry 1983), combine the ascription of perceptual content with a claim about causal connection. A state σ is an NI-perception of state τ in natural-kind v in world w iff there is a state χ of such a kind that:

(i) σ represents χ in v,

(ii) σ robustly carries the information that χ,

(iii) τ entails χ, and

(iv) τ is a cause of σ in w.

This definition supports many of the logical properties of NI-perception reports observed by Barwise and Perry.

$$\models NIP(\sigma, \tau) \rightarrow \tau$$

$$\models NIP(\sigma, \neg\tau) \rightarrow \neg NIP(\sigma, \tau)$$

$$\models NIP(\sigma, (\tau \vee \chi)) \leftrightarrow [NIP(\sigma, \tau) \vee NIP(\sigma, \chi)]$$

$$\models NIP(\sigma, \exists x\tau)) \leftrightarrow \exists x[NIP(\sigma, \tau)]$$

Unfortunately, this definition of naked-infinitive perception does not support one implication that seems intuitively unproblematic: simplification of conjunction.

$$\not\models NIP(\sigma, (\tau \wedge \chi)) \leftrightarrow [NIP(\sigma, \tau) \& NIP(\sigma, \chi)]$$

The reason for this failure pertains to the nature of represention rather than to causation. If $\tau \& \chi$ causes σ, then both τ and χ cause σ, but σ may represent $\tau \wedge \chi$ without representing either infon individually. Whether or not it does may depend on the perceptual apparatus of the perceiver. However, we may introduce a modified definition of naked-infinitive perception to plug this gap:

Definition 6 $w \models NIP{*}(\sigma, \tau) \Leftrightarrow \exists \phi_1 \ldots \phi_n (NIP(\sigma, \phi_1 \wedge \ldots NIP(\sigma, \phi_n) \wedge \{\phi_1, \ldots \phi_n\} \models \tau)$

Naked-infinitive perception, as defined by $NIP*$, has all the desired logical properties.

There are many cognitive states that represent states of affairs without carrying the corresponding information. Propositional attitudes, like beliefs, doubts, thoughts, intentions, as well as desires, imaginings, and other mental states. These states derive their contents from their functional connections with one another, and ultimately, with basic representations (perceptions and motor-volitions). An information-based theory of basic representations can help to break the hermeneutic circle of functionalism, but it is only the first small step toward a general theory of content.

References

Asher, Nicholas, and Michael Morreau. 1991. Commonsense entailment: a modal theory of nonmonotonic reasoning. In *IJCAI-91: Proceedings of the Eleventh International Joint Conference on Artificial Intelligence.* San Mateo, California: Morgan Kaufmann.

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes.* Cambridge, Mass.: MIT Press.

Dretske, Fred I. 1981. *Knowledge and the Flow of Information.* Cambridge, Mass.: MIT Press.

Fodor. Jerry A. 1990. *A Theory of Content and Other Essays.* Cambridge, Mass.: MIT Press.

Gaifman, Haim. 1986. A Theory of Higher-Order Probabilities. In *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1986 Conference*, ed. Joseph Y. Halpern. Los Altos, California: Morgan Kaufmann.

Lehmann, Daniel, and Menachem Magidor. 1992. What does a Conditional Knowledge Base Entail? *Artificial Intelligence* 55:1–60.

Lewis, David. 1973. *Counterfactuals.* Oxford: Blackwell.

Mackie, J. L. 1965. Causes and Conditions. *American Philosophical Quarterly* 2:245–264.

Millikan, Ruth. 1984. *Language, Thought and Other Biological Categories.* Cambridge, Mass.: MIT Press.

Pearl, Judea. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, Calif.: Morgan Kaufmann.

Skyrms, Brian. 1980. Higher Order Degrees of Belief In *Prospects for Pragmatism*, ed. D. H. Mellor. Cambridge: Cambridge University Press.

Spohn, Wolfgang. 1988. Ordinal Conditional Functions: A Dynamic Theory of Epistemic States. In *Causation in Decision, Belief Change and Statistics, II.* Amsterdam: Kluwer Academic.

Stalnaker, Robert. 1986. A Theory of Conditionals. In *Studies in Logical Theory.* American Philosophical Quarterly Monograph Series, No. 2. Oxford: Blackwell.

23

Bridging Situations and NPI Licensing

IK-HWAN LEE

Introduction

In this paper I will present an account of the licensing phenomena of the so-called Non-assertive Polarity Sensitive Items (NPIs, henceforth) within the framework of Situation Semantics (Barwise & Perry 1983, Cooper 1988, and Cooper & Kamp 1991). In particular, I will propose the notion of 'Bridging Situation', which will bridge the licensing of NPIs. For this purpose this paper is organized as follows: In Section 1, I will discuss the 'bridging phenomena' which have been observed in the linguistic literature. In Section 2, the bridging phenomena will be reanalyzed in terms of the situation theory. In Section 3 I will discuss the differences among bridging situation, resource situation, and constraints. The problem of NPI licensing will be discussed in Section 4 In Section 5, I will consider Progovac's (1993) problems and suggest a tentative solution. A further application of the bridging theory will be suggested in Section 6. Section 7 will summarize the arguments and conclude the paper.

This is a revised version of the paper presented at the 4th Conference on Situation Semantics, Moraga, Ca, USA, on June 13–15, 1994 From the initial stage of this paper, many scholars have provided me with helpful comments at various occasions They are Suk-Jin Chang, Hi-Ja Chung, Jae-Woong Choi, Robin Cooper, Young-Se Kang, Keio Komatsu, Susumu Kubo, Chungmin Lee, Kiyong Lee, Young-hern Lee, Byung-Soo Park, Jerry Seligman, Alice ter Meulen I have tried to incorporate their comments into the present version Any remaining defects, however, are all my own This work was supported by a grant from the Korea Research Foundation in 1994

Logic, Language and Computation
Jerry Seligmann and Dag Westerståhl, eds

23.1 Bridging Phenomena

The notion of 'bridging' is borrowed from Clark (1977). He uses the term in accounting for the referential connection between two sentences. Let us look at the following sentences.

(1) Nigel bought <u>a fridge</u>.

(2) <u>The door</u> fell off three weeks later.

(3) The fridge had a door.

(4) Tom <u>went walking</u> at noon.

(5) <u>The park</u> was beautiful.

(6) Tom went walking in the park.

In these sentences, (3) and (6) represent the implicated meaning, which can be constructed by a series of inferences on the basis of what the hearer knows or believes. In particular, (1) and (2) are referentially connected by (3). This will ensure a unique connection. Here the bridging sentence (3) has two functions: First, it helps to locate and identify the referents of the noun phrases, namely, 'a fridge' and 'the door'. Second, it referentially connects the two noun phrases.

Between two sentences, there may be more than one accessible bridges. Look at the sentences in (7)–(8).

(7) Nigel bought a fridge and put it in the caravan.

(8) Three weeks later the door fell off.

Between (7) and (8), we may have two possible bridges, as shown in (9).

(9) a. a fridge — the door: Three weeks later the door of the fridge fell off.

b. the caravan — the door: Three weeks later the door of the caravan fell off.

A natural interpretation prefers the bridge depicted in (9a). Why so? The proposed explanation is that the assignment of reference depends not just on the accessibility of the referent, but also on the accessibility of a context in which the proposition expressed by the speaker is relevant (Blakemore 1992: 76). In order to explain the preferred bridge in (9a) I would like to use the notion of "speaker's perspective", which is similar to, but not exactly the same as, Kuno's (1987: 210) "empathy". This notion is applied as in (10):

(10) a.	In (9a):	Topic Perspective:	P(the fridge) > P(the caravan)
		Surface Perspective:	P(the fridge) > P(the caravan)
b.	In (9b):	Topic Perspective:	P(the fridge) > P(the caravan)
		Surface Perspective:	P(the fridge) < P(the caravan)

where, $P(a) > P(b)$ means that the speaker has a stronger perspective toward a than toward b.

In (10a) there is no logical conflict between two perspective relations defined by the Topic Perspective and the Surface Perspective. Here, the term 'Surface Perspective' is used to cover the hearer's perspective toward his intended (or artificial) interpretation of the connection. In (10b), however, we see a conflict between the two perspective relationships. This fact accounts for the preferred bridge of (9a). This is the speaker's evaluation of the topic continuity of a discourse. We may easily incorporate this fact into the theory of bridging within whatever theoretical framework we adopt.

As a second example, let us examine Wilson & Sperber's (1991[1986]) interesting argument. Look at the dialogue in (11).

(11) a. *John*: Will you have some coffee?
b. *Mary*: Coffee would keep me awake.

From this dialogue a hearer may infer either (12) or (13).

(12) Mary won't have any coffee.

(13) Mary will have some coffee.

We may assume two possible contexts as in (14) and (15).

(14) a. Mary doesn't want to be kept awake.
b. Mary won't have anything that would keep her awake.

(15) a. Mary wants to be kept awake.
b. Mary will have anything on offer that would keep her awake.

Depending on what we assume, we may have one of the two different connections. In a context of (14), (11b) contextually implicates (12), while in a context of (15), (11b) implicates (13), as shown in (16).

(16) a. If we assume (14) as a context: (11b) implicates (12).
b. If we assume (15) as a context: (11b) implicates (13).

Thus, (14) and (15) function as bridging contexts for the implicated conclusions (12) and (13), respectively. In a natural interpretation, (16a) is preferred over (16b). Then, the question is how we can predict this preferred reading. Sperber & Wilson (1991[1986]: 387) observes that the indirect answer (11b) simultaneously refuses the offer of coffee and explains the refusal, thus saving the hearer the time he might have spent speculating on the reasons behind it. They note that an interpretation based on the contextual assumptions in (15) and the conclusion (13) is unlikely to be considered at all. Observing the principle of relevance, a speaker should not prefer the indirect answer (11b) to the direct answer (17).

(17) *Mary*: I want some more coffee.

We may reflect this fact in the bridging theory. For the moment, however, the important point to be noted is that (14) and (15) function as bridging contexts for the implicated conclusion (12) and (13), respectively.

Let us look at a third example. Explaining the conversational maxim of 'Be Relevant' Grice (1991[1968]) constructs the following situation. A is standing by a car without petrol. A is approached by B. The following exchange takes place:

(18) A: I am out of petrol.
B: There is a garage around the corner.

As noted by Wilson & Sperber (1991[1986]: 388), A may have an implicated conclusion as in (20) on the basis of the implicated context given in (19).

(19) If there's a garage round the corner, I can get some petrol there.

(20) I can get some petrol round the corner.

Here again we have a case of bridging context. (19) functions as a bridging context for the implicated conclusion (20).

As noted by Alice ter Meulen (when I presented the idea at Indiana University on June 24, 1994), the three cases discussed so far may seem to be different sorts of examples, but they are similar in that their coherent interpretation necessitates the notion of semantic/pragmatic bridging.

23.2 Bridging in Situation Semantics

In the previous sectin I examined three cases which can be treated in terms of the notion of bridging. Now I will reanalyze these phenomena within the framework of Situation Semantics. I will represent the relationship among sentences in terms of the connection among situations. I will term the bridging context introduced in the previous section the bridging situation.

Bridging situations are represented in terms of infons, which is supported by the situation. Infons in bridging situation will be of two types: conventional infons and conversational infons. Conventional infons are inferred from lexical items and structural features of the sentence uttered in a context. Conversational infons are inferred from the context in which the sentence is used. In the present paper I will mainly talk about the conventional infons.

Now the sentences in Section 1 are represented in situation semantics as follows (examples are repeated with their original numbering):

(1) Nigel bought a fridge.

(2) The door fell off three weeks later.

(3) The fridge had a door.

(21) S_1: Nigel bought a fridge. (for (1))
$S_1 \models \ll buys,\ n,\ f_\imath,\ l_1,\ t_1,\ 1 \gg$

(22) S_2: The door fell off three weeks later.

$S_2 \models \ll \textit{falls-off},\ d_\imath,\ l_2,\ t_2,\ 1 \gg$

(23) S_B: The fridge has a door.

$S_B \models \ll \textit{has},\ f_\imath,\ d_\imath,\ l_B,\ t_B,\ 1 \gg$

[Here, S_B = 'Bridging Situation'; $f_\imath$, $d_\imath$ = uniquely identified fridge and its door; $\models$ = supports]

Now the relation between (21) and (22) is defined in terms of the implicature-relation as in (24).

(24) $[S_1 \wedge S_B] \twoheadrightarrow S_2$, where $\twoheadrightarrow$ means 'implicates'.

Concerning the phenomenon observed with (7), (8), and (9) [repeated with their original numbering], in order to explain the best bridge of the two possible connections I incorporate the perspective theory in the representation of the bridging situation, as shown in (25).

(7) Nigel bought <u>a fridge</u> and put it in <u>the caravan</u>.

(8) Three weeks later <u>the door</u> fell off.

Between (7) and (8), we may have one of the following two possible bridges.

(9) a. a fridge — the door: Three weeks later the door of the fridge fell off.
b. the caravan — the door: Three weeks later the door of the caravan fell off.

(25) S_B: The fridge had a door.
Topic Perspective: P(the fridge) > P(the caravan)
$S_B \models \ll \textit{has},\ f,\ d,\ l_B,\ t_B,\ 1 \gg$
$S_B \models \ll >$, TP(the fridge), TP(the caravan), $l_B,\ t_B, 1 \gg$

Here the first bridging infon is a conventional one and the second infon is a conversational one. The bridging situation (25) helps to connect 'the door' in (8) to the phrase 'a fridge' in (7). Here, the symbol '>' shows the preference degree, and TP means the topic perspective.

Now look at the second case. Here, to connect (11b) to (12), (14) will be postulated as the required bridging situation, as in (26), in which only relevant points are depicted.

(11) a. *John*: Will you have some coffee?
b. *Mary*: Coffee would keep me awake.

(12) Mary won't have any coffee.

(14) a. Mary doesn't want to be kept awake.
b. Mary won't have anything that would keep her awake.

(26) $S_B \models \ll$ *wants*, m, *to-be-awake*, l_B, t_B, $0 \gg$
$S_B \models \ll$ *wants*, m, *to-have-anything-that-would-keep-her-awake*, l_B, t_B, $0 \gg$

The situation representing (11b), together with the bridging situation in (14), i.e. (26), will implicate the situation of (12). Notice that the relationship between (11b) and (12), bridged by the bridging situation, is an implicature, not an entailment.

A similar account can be easily provided for the third case, i.e., examples given in (18)–(19), too.

23.3 Bridging Situation, Resource Situation, and Constraints

One may note that the resource situation introduced in the situation semantics literature may function as the bridging situation postulated in the previous section. Resource situations can become available for exploitation in various ways (Barwise & Perry 1983: 36, Devlin 1991: 227). The function of resource situations, however, is to delimit the domain of indviduals, so that expressions such as definite descriptions may have uniquely identified referents. Suppose John says (27).

(27) The man I saw running yesterday is at the door.

The utterance situation of (27) is (28) (Devlin 1991: 226).

(28) $u \models \ll$ *says*, *John*, ϕ, l, t, $1 \gg$
$\wedge \ll$ *refers-to*, *John*, *THE-MAN*, M, l, t, $1 \gg$
$\wedge \ll$ *refers-to*, *John*, *THE-DOOR*, D, l, t, $1 \gg$

$\phi =$ *THE MAN I SAW RUNNING YESTERDAY IS AT THE DOOR*,
M = a man that is fixed by u; D = a door fixed by u.
In addition to the utterance situation, we may think of another one, a resource situation, r, which John makes use of in establishing that M is the unique man who has the property of 'running' (for some appropriate values of l', t') and who has the property of being at the door. The resource situation is represented as in (29).

(29) $r \models \ll$ *runs*, M, l', t', $1 \gg$

Here the resource situation is used to identify the referent of the definite description and to establish its uniqueness. There is a possibility that this type of resource situation may be extended to accommodate the bridging phenomena.

In general, a bridging situation is to be distinguished from the resource situation (Barwise & Perry 1983), which is postulated in the literature mainly for the purpose of referential identification of definite discriptions or limitation of the referential domain. Specifically, resource situation was postulated for the expressed meaning while bridging situation is for the implicated meaning, which is mainly obtained from the semantic/pragmatic inferences. Resoruce situations were assumed to be different among discourse participants, while a common bridging situation is shared by the discourse participants. Thus, I just postulate the bridging situation for an efficient account of the problem of semantic/pragmatic inferences.

Regarding the implicature relation in (24) (repeated here), one may observe that the same effect can be achieved by the 'involving' relation of the Constraints in Situation semantics.

(24) $[S_1 \wedge S_B] \twoheadrightarrow S_2$, where $\twoheadrightarrow$ means 'implicates'.

In order to explain the meaning of a sentence such as (30), situation semantics defines an 'involve' relation (Devlin 1991: 88; Barwise 1993).

(30) Smoke means fire.

It is claimed that situations where there is smoke are linked to situations where there is a fire. This linkage is accounted for by the notion of 'constraint'. The two situations can be represented as in (31).

(31) a. $S_0 = [\, s_0' \mid s_0' \models \ll \textit{smoke-present},\ l',\ t',\ 1 \gg]$
b. $S_1 = [\, s_1' \mid s_1' \models \ll \textit{fire-present},\ l',\ t',\ 1 \gg]$

The constraint is denoted by the expression in (32).

(32) $S_0 \Rightarrow S$ $(= \ll \textit{involves},\ S_0,\ S_1,\ 1 \gg$; i.e., S_0 involves $S_1)$

This type of constraint is called a nomic constraint, which corresponds to some natural law. Other sentence types require different constraints. Let us look at the examples in (33).

(33) a. Kissing means touching.
b. The ringing bell means class is over.

For the meaning of the sentence (33a), we need a reflexive constraint, while a conventional constraint is needed for (33b).

Notice that the bridging phenomena are highly context-dependent inferences of semantic and pragmatic nature. 'Involve' relations, on the other hand, are stipulated constraints. The bridging relation cannot be defined as 'nomic', or as 'reflexive', or as 'conventional'. Once the bridging situation is added to the described situation of the first sentence, then the conjunction implicates the described situation of the second sentence. This is not the type of 'involve' relation defined in terms of an entailment relation. Therefore, it is reasonable to postulate the bridging situation as a significant tool of semantics and pragmatics of natural language.

23.4 Non-assertive Contexts

NPI phenomena may look different from examples we have discussed. But we can assume that the use and understanding of NPI constructions require similar inference steps. Thus, I attempt to explain the NPI phenomena by the notion of bridging infons in situation semantics.

Following Quirk et al. (1985: 83, 780), by 'non-assertive' contexts I mean negative context and half-negative-half-positive ones, which may be classified as in (34).

(34) Non-assertive contexts

A. Morphologically marked (i.e., lexical)

1. Explicit Negation
2. Implicit Negation:

(i). Incomplete negation: hardly, little, few, least, seldom, ...

Ex. John hardly undersood any of the point.

(ii). Implied negation: before, fail, prevent, reluctant, surprised,

difficult, hard, ...

Ex. I was surprised/sorry that he ever said anything.

B. Morphologically unmarked (i.e., non-lexical)

Ex. If John subscribes to any newspaper, he gets well informed.

Does John subscribe to any newspaper?

Everyone who has any matches is happy.

(Kadmon & Landman 1993: 370)

In this paper I am particularly concerned to provide a proper account of the non-assertive cases given in A.2 (implicit negation) and B (morphologically unmarked). For this purpose, I would like to use the mood indicator introduced by Grice (1991[1968]). He defines the utterance meaning as in (35).

(35) By uttering x, U meant that $^*p$.

In (35) the symbol '*' means a dummy mood indicator, i.e., the unspecified mood indicator, and U the utterer. It may be replaced by one of the specific mood indicators as shown in (36).

(36) a. $\vdash$ (indicative/assertive)

b. ! (imperative)

To distinguish grammatical forms of sentences I would like to add three more as in (36c, d, e).

(36) c. ? (interrogative)

d. $\neg$ (negative)

e. $\nvdash$ (non-assertive)

A p with the symbol '*' in (35) is an unspecified proposition, which may be replaced by one of the specific mood indicators given in (36). This replacement is triggered by the factors given in (34). Actually, (36c) and (36d) are special cases of (36e) because the interrogative and negative moods belong to the non-assertive category. For the expository purpose, however, I would like to use them interchangeably. This subdivision is useful in categorizing *indicative conditionals and relative clauses with universally quantified* head. The examples in (37) show the details of this device.

(37) a. John understood <u>some</u> portion of Mary's story.
b. John <u>hardly</u> understood <u>any</u> portion of Mary's story.
c. $\not\models$ John understood a portion of Mary's story.

In (37) the word 'hardly' signals that the sentence internal context is incomplete negation, which is represented as in (37c).
(37c) is the negative counterpart of (37a), which is an assertive sentence. Thanks to the word 'hardly' we may postulate (37c) as a context with negative force. This negative context licenses the NPI 'any' in (37b). Notice that 'any' may not be licensed in (37a). Here, I take (37c) to be a bridging situation, which connects two situations denoted by (37a) and (37b). Details aside, the idea is represented as in (38).

(38) a. S_1: John understood <u>some</u> portion of Mary's story.
$S_1 \models \ll \exists\, x[\textit{P-o-M's-S}(x) \wedge \textit{understands},\ j,\ x,\ l_1,\ t_1, 1] \gg$
b. S_2: John <u>hardly</u> understood <u>any</u> portion of Mary's story.
$S_2 \models$ [John <u>hardly</u> understood <u>any</u> protion of Mary's story]
c. $S_B \models \ll \not\models \exists x[\textit{P-o-M's-S}(x) \wedge \textit{understands}, j,\ x],\ l_B,\ t_B,\ 1 \gg$

An abbreviated representation for (38):

John <u>hardly</u> understood <u>any</u> portion of Mary's story.
ϕ: John understood <u>some</u> portion of Mary's story.
$S_1 \models \ll \phi,\ l_1,\ t_1,\ 1 \gg$
$S_B \models \ll \textit{implicates},\ \textit{hardly-}\phi,\ \not\models \phi,\ l_B,\ t_B,\ 1 \gg$
$S_2 \models \ll \textit{used},\ \textit{NPI(any)},\ \textit{in hardly-}\phi,\ l_2,\ t_2,\ 1 \gg$
$u \models \ll [S_1 \wedge S_B] \twoheadrightarrow S_2 \gg$, where $\twoheadrightarrow$ means 'implicates'.

The implicature relation of the situations is defined as in (39).

(39) $[S_1 \wedge S_B] \twoheadrightarrow S_2$, where $\twoheadrightarrow$ means 'implicates'.

S_2 in (38) is a rough representation. To give an exact logical formulation of the situation, we need the exact semantic representation of the word 'hardly'. This is not directly relevant here. The important point to be noted is that we can naturally explain the licensing of 'any' in (37c), i.e, S_2 in (38) by postulating a bridging situation (S_B). Note that this bridging situation connects (37a) and (37b). That is, 'hardly' triggers the bridging situation (38c), which licenses the NPI in S_2, i.e. (37b). Here, (38c) may be taken to be an utterance situation in which the NPI is licensed.

A similar account can be provided for the sentences in (40) by postulating bridging situations denoted by the semi-logical forms in (41).

(40) a. John died <u>before</u> he finished any thesis.
b. Does John subscribe to any newspaper?

(41) a. $\neg$ [John finished a thesis]
b. $\not\vdash$[John subscribes to a newspaper]
(or ?[John subscribes to a newspaper])
(N.B.: $\neg$ and ? are two examples of the symbol $\not\vdash$.)

Abbreviated representations for (40) and (41):

(40a & 41a) John died <u>before</u> he finished any thesis.
ϕ: John finished a thesis.
$S_1 \models \ll \phi,\ l_1,\ t_1,\ 1 \gg$
$S_B \models \ll \textit{implicates},\ \textit{before-}\phi,\ \not\vdash \phi,\ l_B,\ t_B,\ 1 \gg$
$S_2 \models \ll \textit{used},\ \textit{NPI(any)},\ \textit{in before-}\phi,\ l_2,\ t_2,\ 1 \gg$

(40b & 41b) Does John subscribe to <u>any</u> newspaper?
ϕ: John subscribes to <u>a</u> paper.
$S_1 \models \ll \phi,\ l_1,\ t_1,\ 1 \gg$
$S_B \models \ll \textit{implicates},\ ?\phi,\ \not\vdash \phi,\ l_B,\ t_B,\ 1 \gg$
$S_2 \models \ll \textit{used},\ \textit{NPI(any)},\ \textit{in}\ ?\phi,\ l_2,\ t_2,\ 1 \gg$

For the pair (40a) and (41a), the explanation for (38) applies with minor revisions. What is important is that the implicit neg-expression 'hardly' and 'before' triggers a bridging situation.

The pair (40b) and (41b), however, requires another refinement. As noted above, the symbol '$\not\vdash$ ' represents the 'non-assertive' force. However, this symbol will be used to designate indicative conditionals, adversative predicates such as 'doubt', comparatives, relatives with a universally quantified head, etc. As noted by Susumu Kubo (personal communication) even in (41a) we may use the same symbol. But I used two different symbols for a sub-division of the non-assertive contexts. This is roughly equivalent to Progovac's (1993: 163) minus value of the Polarity Operator (Op).

Now, we are ready to consider other relevant examples, which are cited from various sources. The representation will be simplified just to indicate the relevant points and for the brevity of presentation.

23.4.1 Conditionals

(42) a. If anyone can do that, we will reward that person heavily. (Progovac (11a))
b. S_B: $\not\vdash$ [One can do that] (Neither positive nor negative)

(43) If a sentence has the form of [If ϕ, then ψ], then the following situation holds: $\not\vdash$ [ϕ].

For (42a) a bridging situation such as (42b) can be postulated on the basis of a general principle given in (43). This will account for the appearance of NPI in the antecedent of (42a). An abbreviated representation for (42a) may be given as follows:

If anyone can do that, we will reward that person heavily.
ϕ: One can do that.
$S_1 \models \ll \phi,\ l_1,\ t_1,\ 1 \gg$
$S_B \models \ll \textit{implicates},\ \textit{if-}\phi, \not\models \phi,\ l_B,\ t_B,\ 1 \gg$
$S_2 \models \ll \textit{used},\ \textit{NPI(any)},\ \textit{in if-}\phi,\ l_2,\ t_2,\ 1 \gg$

23.4.2 Relatives with Universally Quantified Head

(44) Everyone who ever reads books will be pleased with this proposal.

I would like to assume that (44) can be paraphrased as in (45).

(45) If one ever reads books he will be pleased with this proposal.

In other words, I interpret the relative clause in (44) in terms of its conditional counterpart. This way we can account for the licensing of NPI in (44) by using the method introduced in (42)–(43). This method conforms to the traditional logic, which translates a universal quantifier into a conditional form. For the case of (44), we may postulate a redundant bridge, which will not be harmful at all in the semantics of NPI licensing. Accordingly, the bridge for (44) will be stipulated as in (46).

(46) a. Everyone who ever reads books will be pleased with this proposal. (= 44)
b. If one ever reads books he will be pleased with this proposal. (= 45)
c. S_B: $\not\models$ [One reads books]

In (46), we do not actually need to worry about the relationship between a. and b. since they are logically equivalent. However, b. is necessary to induce c. on the basis of (43).

23.4.3 Interrogatives

(47) Does Bill trust anyone?

(48) When does Bill trust anyone?

Interrogatives are very simple. Both (47) and (48) will have a similar bridging situation such as the one in (49).

(49) S_B: $\not\models$ [Bill trusts someone] (or ? [Bill trusts someone])

Let us now return to cases where a bridging situation with negative force is required.

23.4.4 Only

(50) Only John has ever been there.

(51) S_B: $\forall x(\neg[x = \textit{John}] \rightarrow \neg(P(x)))$

The constructions with such expressions as 'at most NP', 'exactly NP', etc. can be explained by assuming that they contain the bridging situations equivalent to that of 'only NP'.

23.4.5 Adversative Predicates

(52) Mary was amazed that there was any food left. (Linebarger 1987)

(53) Mary expected that there wouldn't be any food left.

(54) $S_B = \mathit{expect}(m, \neg[\phi])$,
where $[\phi]$ = There would be some food left.

(55) I'm sorry that anybody hates me.

(56) I want nobody to hate me.

(57) $S_B = \mathit{want}(\mathit{Speaker}, \neg[\phi])$,
where $[\phi]$ = Somebody hates me.

In (52), due to the presence of the predicate 'amazed', we have (53) as an intuitive reading. This functions as a bridging situation, the simplified representation of which is given in (54). Exactly the same phenomenon holds for (55).

Considering the analyses given so far, I provide a general principle as in (58).

(58) An NPI is licensed in the following context:

a. an explicit negative sentence, or
b. a sentence that may have a bridging situation with non-assertive mood indicator (i.e., $\neg$, ?, $\not\vdash$)

23.5 Bridging Solution to Progovac's Problems

Now I would like to discuss problems raised by Progovac (1993) and suggest tentative solutions in terms of bridging situation.

First, Progovac (1993: 176) notes that in English the sentence initial 'only' licenses NPI and PPI (Positive Polarity Item) as well as shown in (59).

(59) a. Only Mary showed <u>some</u> respect for the visitors.
b. Only Mary showed <u>any</u> respect for the visitors.

She appeals to some syntactic operation for a solution. In my bridging approach, this problem can be easily resolved. Although it is not uncontroversal, it is a common practice to analyze the structure with the focusing delimiter 'only' into two parts, namely Assertion (Extension) and PresupposAiton (Conventional Implicature) (Karttunen & Peters 1979; Ik-Hwan Lee 1977; Horn 1989; Atlas 1991). In this analysis the sentence (60) is represented as in (61).

(60) Only Mary likes John.

(61) a. Presupposition (Implicature): Mary likes John.
b. Assertion (Extension): No one other than Mary likes John.

We may take these two types of meaning to be two different bridging situations. Licensing of PPI and NPI in (59a) and (59b), respectively, are represented as in (62).

(62) a. Only Mary...: Presupposition Reading: (59a) is licensed.
b. Only Mary...: Assertion Reading: (59b) is licensed.

That is, in (62a) the presupposition reading functions as the bridging situation, while in (62b) the bridging situation is the assertion reading. As noted by Jerry Seligman (when I presented this idea at Indiana University on June 24, 1994), it seems true that (59a) still had some sort of negative force. But what I observe here is that the speaker, by uttering this sentence, intends to deliver the presupposition part of meaning, not the negative assertion part. At first sight this treatment may sound odd. But in actual language use, this kind of practice is not infrequent. Look at the question-answer pairs in (63)–(64).

(63) a. Whom does Mary like?
b. Mary likes Bill.

(64) a. Whom does Mary like?
b. No one.

It is generally assumed that a WH-question presupposes that there is a true answer to the question. In particular, the question in (63a) presupposes that 'there is someone whom Mary likes'. In (63) this presupposition is observed and (63b) is given as a true answer. On the other hand, in (64) the presupposition itself is denied (Hausser & Zaefferer 1979: 348). This phenomenon shows that a given sentence may be perceived in two ways, i.e., either as its assertion or as its presupposition.

Second, Progovac (1993) notes that the delimiter 'only' seems to license NPIs only from the front sentential position. She gives the following sentences.

(65) $[_{IP}$ Only Mary showed any respect for the visitors.$]$

(66) $[_{CP}$ Only to his girlfriend $[_{C'}$ did $[_{IP}$ John give any flowers.$]]]$

(67) $[_{CP}$ Only last year $[_{C'}$ did $[_{IP}$ John get any grey hairs.$]]]$

(68) ?*John gave only his girlfriend any flowers.

(69) ?*John told only Mary about any books.

As a syntactic explanation of the phenomenon, she argues that 'only' can become a licenser only if it raises to Comp, implying that it is in Comp even in (65).

We may look at the problem from a simple point of view. For example, in (68) the subject 'John' may be regarded to be the topic. The phrase 'his girlfriend' is the focus of the particle 'only'. What is the sentence talking about? If we think the subject 'John' to be the topic (or a near topic), then the sentence is about John. However, if we regard the phrase 'only his girlfriend' to be a focal point, then the sentence is about this phrase, too. We may incorporate the notion of 'aboutness' (Atlas 1991) into the bridging situation. It is difficult to postulate two different phrases in a sentence as objects of 'aboutness'. This seems to be another case of Perspective conflict, and this explains the uncomfortable acceptance of sentences such as (68) and (69).

Her third problem concerns the adverbial 'rarely' as in (70).

(70) John <u>rarely</u> speaks to <u>anyone</u>.

Her problem is the movement of this adverbial to Comp. Again, this seems to be a syntactic problem. Semantically, (70) can be treated in the same way as the expression 'hardly' was accounted for in (37)–(38). As before, the expression 'rarely' triggers a bridging situation with negative force.

Thus, the bridging situation approach seems to provide a satisfactory account for various problems related to NPIs.

23.6 A Further Application

The bridging theory can be applied for an appropriate account of some aspectual connections of sentences. I would like to examine some examples here. Smith (1986) discusses somewhat non-standard aspectual sentences such as those given in (71).

(71) a. John was really liking the play.
b. Amy is resembling her great-uncle.
c. That cake is looking done.

These sentences talk about states, or stative situations. It is argued that they present states as if they were events. On the basis of these and similar other examples, Smith concludes that depending on the speaker's perspective or viewpoint an actual situation may be represented differently. This is termed 'the speaker's aspectual perspective'.

In my analysis (Lee 1994), this perspective is represented as 'the speaker's perspective', which is included in 'bridging situation'. This bridging situation bridges two situations, namely the actual situation and the one manipulated by the speaker. For example, (71a) is represented as in (72).

(72) a. Actual situation: stative situation (John likes the play)
b. Bridging situation: Speaker's perspective (Progressive aspect)
c. Ideal situation: activity situation (John is liking the play)

An abbreviated representation:

ϕ: John liked the play.
$S_1 \models \ll state,\ \phi,\ l_1,\ t_1,\ 1 \gg$
$S_B \models \ll [\mathrm{P}(state) < \mathrm{P}(event)],\ l_B,\ t_B,\ 1 \gg$
$S_2 \models \ll event,\ \phi,\ l_2,\ t_2,\ 1 \gg$

In this way, by using the notion of bridging situation, we can provide a proper account of the licensing of the somewhat non-standard aspectual sentences.

23.7 Summary and Conclusion

In this paper an attempt was made to provide a proper account of the licensing phenomena of the so-called NPIs (Non-assertive Polarity Items) within the framework of Situation Semantics. For this purpose I introduced the notion of bridging situation, which is a situation semantics version of Clark's (1977) notion of bridging. I discussed several different aspects of bridging phenomena in Section 1, which were reanalyzed in terms of situation semantics in Section 2. In Section 3 the notion of bridging situation was justified as a theoretical tool in situation semantics. In Section 4 I discussed various cases of NPI licensing and provided a new analysis by using bridging situations. Progovac's (1993) problems were investigated and tentative solutions were suggested in Section 5. In Section 6, I exemplified a further application of the theory.

To conclude, this paper has shown that the notion of bridging situation is helpful in providing a proper treatment of NPI licensing and other linguistic aspects.

References

Atlas, Jay David. 1991. Topic/Comment, Presupposition, Logical Form and Focus Stress Implicatures: the Case of Focal Particles 'Only' and 'Also'. *Journal of Semantics* 8:127–148.

Barwise, Jon. 1993. Constraints, Channels, and the Flow of Information. In *Situation Theory and Its Applications, vol.3.* CSLI Lecture Notes. Stanford: CSLI Publications.

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes.* Cambridge, MA: MIT Press.

Blakemore, Diane. 1991. *Understanding Utterance: An Introduction to Pragmatics.* Oxford: Blackwell.

Clark, Herbert. 1977. Bridging. In *Thinking: Readings in Cognitive Science*, ed. P. Johnson-Laird and P. Watson. Cambridge: Cambridge University Press.

Cooper, Robin. 1988. Facts in Situation Theory: Representation, Psychology, or Reality? In *Mental Representations*, ed. Ruth Kempson. Cambridge: Cambridge University Press.

Cooper, Robin, and Hans Kamp. 1991. Negation in Situation Semantics and Discourse Representation. In *Situation Theory and Its Applications, vol.2*, ed. Jon Barwise, Mark Gawron, Gordon Plotkin, and Syun Tutiya. Stanford: CSLI Publications.

Devlin, Keith. 1991. *Logic and Information*. Cambridge: Cambridge University Press.

Davies, Steven. 1991, ed. *Pragmatics: A Reader*. Oxford: Oxford University Press.

Grice, H. P. 1991. Utterer's Meaning, Sentence-Meaning, and Word-Meaning. In *Pragmatics: A Reader*, ed. Steven Davies. 65–76. Oxford: Oxford University Press.

Hausser, Roland, and Ditmar Zaefferer. 1979. Questions and Answers, in a Context-dependent Montague Grammar. In *Formal Semantics and Pragmatics for Natural Language*, ed. F. Guenthner and S.J. Schmidt. Dordrecht: D. Reidel.

Horn, Larry. 1989. *A Natural History of Negation*. Chicago: University of Chicago Press.

Jhee, In-Young. 1993. *An Analysis of Negation in Situation Semantics*. Yonsei University: Ph.D. Dissertation.

Kadmon, Nirit, and Fred Landman. 1993. Any. *Linguistics and Philosophy* 16:353–422.

Karttunen, Lauri, and Peters Stanley. 1979. Conventional Implicature. In *Syntax and Semantics, 11.ed.*, ed. Choon-Kyu Oh. New York· Academic Press.

Kuno, Susumu. 1987. *Fuctional Syntax*. Chicago: The University of Chicago Press.

Ladusaw, William. 1979. *Polarity Sensitivity as Inherent Scope Relations*. Austin, TX: Ph.D. Dissertation.

Lee, Ik-Hwan. 1977. Syntax and Semantics of Korean Delimiters. *CLS* 13:302–312.

Lee, Ik-Hwan. 1994. Interpretation of Temporal Connections in Situation Semantics. A paper presented at the 6th Asian Conference on Information, Languages, and Computation, August 10–12, 1994, Kyoto University. Kyoto.

Linebarger, Marcia. 1987. Negative Polarity and Grammatical Representation. *Linguistics and Philosophy* 10:325–87.

Progovac, Ljiljana. 1993. Negative Polarity: Entailment and Binding. *Linguistics and Philosophy* 16:149–180.

Smith, Carlota. 1986. A Speaker-based View of Aspect. *Linguistics and Philosophy* 9:97–115.

Wilson, Deidre, and Dan Sperber. 1991. Inference and Implicature. In *Pragmatics: A Reader*, ed. Steven Davies. 377–393. Oxford: Oxford U Press.

24

A Diagrammatic Inference System for Geometry

ISABEL LUENGO

Introduction

The main claim of this paper is that diagrams can be used in geometry proofs in essential ways. We will present a visual inference system for geometry describing its syntax and semantics. The system contains both inference rules and construction rules. To validate the construction rules typically involved in geometry reasoning a new concept of consequence is needed. We will define this new concept (we will call it *geometric consequence*) and we will use it to prove the system to be sound.

24.1 Syntax

24.1.1 Diagrammatic Objects

There are two basic kinds of diagrammatic objects: primitive and derived. The primitive diagrammatic objects are not defined. The derived diagrammatic objects are defined in terms of the primitive objects.

24.1.1.1 Primitive Objects

In this particular system there are six disjoint kinds of primitive objects. We could add more kinds (for instance, to represent circles) depending on our needs.

1. Boxes
   A box is any rectangle whose edges are dashed lines.
2. Points*
   I use the asterisk to distinguish points* from their spatial counter-

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

parts. Points* are not mathematical, abstract entities, but very concrete, physical objects. They are small dots.

3. Lines*
   Again, the asterisk is used to distinguish between the physical objects and the mathematical entites. A line* is a straight line.
4. Congruence indicators
   A congruence indicator is an arc with n transversal bars on it, for some $n \geq 1$. There are infinitely many types of congruence indicators. Congruence indicators γ and δ belong to the same type if and only if they have the same number of bars.
5. Parallel indicators
   A parallel indicator is a sequence of n integral symbols, for some $n \geq 1$. There are infinitely many types of parallel indicators. Parallel indicators γ and δ belong to the same type if and only if they have the same number of integral symbols.
6. Labels
   There are two different kinds of labels: point labels and line labels.
   Point Labels: $A, B, C, A', A'', A'''... A_1, A_2, A_3...$
   Line Labels: $l, m, n, l', l'', l'''...l_1, l_2, l_3...$
   The set of labels is called $\mathcal{L}$.

24.1.1.2 Diagrammatic Relations

There are only six relations among primitive diagrammatic objects that are relevant in this system:

1. *In** $\subseteq$ diagrammatic objects $\times$ boxes
   A diagrammatic object x is *in** a box y iff every point of x is within the perimeter of y.
2. *Between** $\subseteq$ points* $\times$ (points* $\times$ points*)
   Point* x is *between* $*\ y$ and z iff there is a line* u such that x, y, z are on* u and x is between y and z.
3. *On** $\subseteq$ points* $\times$ lines*
   Point* x is on* line* y iff x and y intersect.
4. *On** $\subseteq$ parallel indicators $\times$ lines*
   Parallel indicator x is on* line* y iff x and y intersect.
5. *On** $\subseteq$ congruence indicators $\times$ (points* $\times$ points*)
   If y and z are points* congruence indicator x is on* $\langle y, z\rangle$ if and only if one end of x intersects with y and the other end intersects with z.
6. *On** $\subseteq$ congruence indicators $\times$ (lines* $\times$ lines*)
   If y and z are lines* congruence indicator x is on* $\langle y, z\rangle$ if and only

if one end of x intersects with y and the other end intersects with z.

Notice that four of the relations have the same name: $On*$. The context will make clear which one we are talking about.

Only these six relations will be representing relations – all other relations among diagrammatic objects will be considered accidental. For instance, if lines* x and y intersect but they don't have any common points* on* them, the fact that they intersect will be accidental, and no interpretation will be given to it, as we will see later when we study the semantics of the system. However, the decision as to what relations among diagrammatic objects represent is somewhat arbitrary. We could have a different inference system in which the fact that two lines* intersect would be considered a diagrammatic relation.

24.1.1.3 Derived Objects

There are four kinds of derived objects:

1. Segments*
   A segment* consists of two distinct points* on* a line* l and the part of l that lies between them. It can be modeled as a pair $\langle l, s\rangle$, where l is a line* and s is a set of two distinct points* on* l.
   The segment* defined by points* A and B is called AB or BA.
   Congruence indicator x is on* segment* AB if and only if x is on* $\langle A, B\rangle$ and A and B are on the same line.[1]

1. Angles*
   An angle* consists of two segments* that have one point* in common.
   The angle* defined by segments* AB and BC is called $\angle ABC$ or $\angle CBA$.
   If B' is between* A and B, then $\angle ABC = \angle AB'C$.
   Congruence indicator x is on* angle* $\angle ABC$ if and only if, if A and B are on* l and B and C are on* m, x is on* $\langle l, m\rangle$.
2. Triangles*
   Defined by three different segments* such that each one has a point* in common with each of the other two, but there is no point* in common to the three of them.
   The triangle* defined by segments* AB, BC, and CA is called $\triangle ABC$.
3. Side of a line*
   Points* x and y are on the same side of line* u with respect to

[1]Remember we have defined $on*$ as a relation between a congruence indicator and a pair of points.

point* z iff x, y and z are on* u and x is between* y and z, or y is between* x and z.

The relation between x and y of being on the same side of z is an equivalence relation on points* on* a line*. Each equivalence class is called *a side of* u *with respect to* z.

24.1.2 Well-Formed Diagrams

Every finite combination of diagrammatic objects is a diagram. But not all diagrams are well-formed diagrams. $\mathcal{D}$ is the set of well-formed diagrams.

Definition 1 A diagram D is *well-formed* if and only if:

1. there is one and only one box in D and all the other diagrammatic objects of D are in* it,
2. every congruence indicator in D is on* either a segment* or an angle*, but not both,
3. every parallel indicator in D is on* a line*,
4. every point* and line* has at least one label. That is, there is a total function *Label* from the set of points* and lines* of D to $\mathcal{P}(\mathcal{L}) - \{\emptyset\}$, such that:
   a. if y is a point*, then if $x \in Label(y)$, x is a point label, and
   b. if y is a line*, then if $x \in Label(y)$, x is a line label.

24.1.3 Conditions on Readability of Diagrams

Well-formed diagrams need to meet some pragmatic conditions to make them readable. A diagram does not have to meet these conditions in order to be well-formed, but if it does not meet them then it will be practically of no use in our proofs. One condition is that each point* label should be clearly closer to the point* it labels than to any other point*, and the same thing applies to line* labels.

Another obvious condition is that all the diagrammatic objects should be large enough for the human eye to see them. When a well formed diagram is not readable we can redraw it (make a copy of it) to insure that these pragmatic concerns can be met.

24.1.4 Diagrams as Equivalence Classes

Definition 2 Wfd. D is an *extension* of E ($E \subseteq D$) iff there is a 1-1 total function g from the diagrammatic objects of E to the diagrammatic objects of D such that:

i. for all $x, g(x)$ is of the same kind as x,
ii. x is on* y, if and only if $g(x)$ on* $g(y)$[2]

[2]Notice that this condition includes the four different notions of on* we have discussed.

iii. x is between* y and z if and only if $g(x)$ is between* $g(y)$ and $g(z)$,
iv. $Label(x) = Label(g(x))$

Definition 3 Wfd. E is a *copy* of D iff $E \subseteq D$ and $D \subseteq E$.

Since *being a copy of* is an equivalence relation, all the diagrams that are a copy of a given diagram D form an equivalence class. From now on by D I will mean the equivalence class of all the diagrams that are a copy of D.

24.2 Semantics

The goal of our diagrammatic system is to represent spatial situations. A spatial situation consists of a set of lines and points on a Euclidean plane where a Euclidean plane is any model of Hilbert's axioms for Euclidean geometry.

Definition 4 $\mathcal{M} = \langle s, f \rangle$ is a *model* of D ($\mathcal{M} \models D$) if and only if s is a set of lines and points on a Euclidean plane and f is a function from the diagrammatic objects of D to the objects of s such that:

1. if $Label(x) \cap Label(y) \neq \emptyset$ then $f(x) = f(y)$,
2. if x is a point* then $f(x)$ is a point,
3. if x is a line* then $f(x)$ is a line,
4. if x is a point* and y is a line* and x is on* y then $f(x)$ lies on $f(y)$,
5. if x is between* y and z, then $f(x)$ is between $f(y)$ and $f(z)$,
6. if xy is a segment*, then $f(xy)$ is the segment defined by $f(x)$ and $f(y)$,
7. if $\angle xyz$ is an angle*, then $f(\angle xyz)$ is the angle defined by $f(xy)$ and $f(yz)$,
8. if Δxyz is a triangle*, then $f(\Delta xyz)$ is the triangle defined by $f(xy)$, $f(yz)$, and $f(xz)$,
9. if c and d are congruence indicators of the same type and x and y are both angles* or both segments*, then if c is on* x and d is on* y, $f(x)$ and $f(y)$ are congruent,
10. if c and d are parallel indicators of the same type and l and l' are both lines*, then if c is on* l and d is on* l', $f(x)$ and $f(y)$ are parallel, and
11. if x and y are on the same side of line* z with respect to point* u, then $f(x)$ and $f(y)$ are on the same side of $f(z)$ with respect to $f(u)$.

What objects in the diagram and relations among diagrammatic objects we are going to take to represent objects and relations in the Euclidean space is in a way the result of an open decision. Thus, in this particular system the fact that two lines* intersect does not imply that there are two abstract lines that intersect, or the fact that two lines* are different does not imply that there are two different lines on the Euclidean plane, but that does not mean that we could not have another diagrammatic system for which those two facts were representing facts, meaning that the interpretation function would assign them a mathematical value.

A diagrammatic system is created with a specific purpose in mind. In the system under consideration the choices we have made are aimed to reflect the common usage of diagrams in informal geometrical proofs as much as possible. However, we are aware that we might need to change some of the conditions the interpretation function has to meet, if they do not prove to contribute to the purpose of the system.

Definition 5 E is a *logical consequence* of D (written, $D \models E$) if and only if, for every $\mathcal{M} \models D$, $\mathcal{M} \models E$.

The notion of logical consequence is not enough to account for the validity of many rules that we need to have in an inference visual system for geometry. For instance, we would like to have in our inference system a Rule of Introduction of a Line* such that, given a diagram D with only two points A, B, would allow us to get a new diagram E identical to D but for the fact that there is a line l in E such that A and B are on it. This rule could be established as follows:

Rule 1 If D has two points* x and y then, if E is an extension of D adding only a line* z and a label for z such that x and y are on* z, then $\{E\}$ is obtainable from D.

Rule 1 should be valid, according to the first of Hilbert's axioms:

> For every two points A, B exists a line a that contains each of the points A, B.

However, we can notice that Rule 1 is not valid in the logical sense, since there is a model $\mathcal{M} \models D$ that it is not a model of E, namely all the models that give values to A and B but not to l.

Thus, we need a new notion of consequence to validate the construction rules typically involved in the informal use of diagrams in geometric proofs.

Definition 6 $\mathcal{M}' = \langle s, f' \rangle$ is an *extension* of $\mathcal{M} = \langle s, f \rangle$ ($\mathcal{M} \supseteq \mathcal{M}'$) iff $f \supseteq f'$.

Definition 7 E is a geometric consequence of D (written, $D \models\!\!\subset E$) if and only if, for any model $\mathcal{M}$ of D that doesn't give values to the diagrammatic objects of E not in D there is a model $\mathcal{N} \supseteq \mathcal{M}$ such that $\mathcal{N} \models E$.

Lemma 1 *If $D \models E$ then $D \models\!\!\subset E$.*

Proof. Based on the fact that each model is an extension of itself. □

We can generalize the notion of geometric consequence to sets of diagrams as follows:

Definition 8 S is a *geometric consequence* of D ($D \mid\!\subset S$) if and only if for each $\mathcal{M} \models D$ that doesn't give values to the diagrammatic objects of S not in D there is a $\mathcal{N} \supseteq \mathcal{M}$ such that $\mathcal{N} \models E$ for some $E \in S$.

24.3 Rules of Transformation

The rules of transformation allow us to obtain a set of diagrams S from a given diagram D.[3] There are two different kinds of rules of transformation corresponding to the two different notions of consequence: inference rules and construction rules.

When we apply an inference rule to D we get a set of diagrams S such that if E is in S then $E \subseteq D$ (erasure rules), or E is the result of copying D adding a new congruence or parallel indicator. Rule 2 is an example of inference rule.

Rule 2 If D is such that

1. there is an indicator of type γ on* x and an indicator of type γ on* y,
2. there is an indicator of type δ on* y, and an indicator of type δ on* z, and
3. x, y and z are the same kind of object,

then if E is an extension of D adding an indicator of type γ on* z, $\{E\}$ is obtainable from D.

When we apply a construction rule to D we get a set of diagrams S such that there is some E in S that is the result of copying D adding a new line* or point*. Rule 1 is an example of a construction rule.

We could have different different diagrammatic inference systems for geometry with different sets of transformation rules, but in all cases both inference rules and construction rules would be necessary.

We notice that only one diagram is the result of the application of Rules 1 and 2. But in some other cases we should get a disjunction of diagrams. We would presumably want a construction rule to correspond to the following axiom of Hilbert:

> If A, B are two points on a line l, and A' is a point on the same or on another line l' then it is always possible to find a point on a given

[3] Intuitively a set of disjunctive, exhaustive cases.

side of the line l' through A' such that the segment AB is congruent or equal to the segment $A'B'$.

If there is one or more points on a given side of l' then there are several diagrams that we could get from the application of the rule; we only need to make sure that the set of diagrams we can obtain exhausts all the possibilities. The rule we need could be estalished as follows:

Rule 3 If D has a point* x on* a line* z and a segment* $x'y'$ with a congruence indicator of type γ on* it, then for each side s in which z is divided by x there is a set S obtainable for D such that $E \in S$ iff E is an extension of D adding:

1. a new point* y on* z (and a label for y) and a congruence indicator of type γ on* segment* xy, or
2. a congruence indicator of type γ for some y on* z.

Definition 9 A set S is *obtainable* from D ($D \rightsquigarrow S$) if and only if there is a rule of transformation R such that S is the result of applying R to D.

24.4 Derivations

Since they typically involve the use of cases exhaustive, geometry proofs that use diagrams cannot have the form of a sequence, like in most linguistic formal systems. In fact they will have a graph-structure.

In order to define the notion of derivation in a diagrammatic inference system for geometry I am going to use some graph terminology borrowed from Barwise and Etchemendy 1990, with some modifications.

Definition 10 A *derivation* $\mathcal{G}$ from D is a rooted, finite, directed graph such that

1. all the nodes are diagrams,
2. the initial node is D,
3. for all D' the set S of children of D' is obtainable from D'

Definition 11 S is *derivable* from D ($D \vdash S$) if and only if there is a sequence $\Gamma =< \mathcal{G}_0, ..., \mathcal{G}_n >$ such that

1. $\mathcal{G}_0 = \{D\}$
2. For each $\mathcal{G}_m (0 \leq m < n)$, $\mathcal{G}_{m+1}$ is like $\mathcal{G}_m$ but for the fact that there is one and only one terminal node in $\mathcal{G}_m$ that has children in $\mathcal{G}_{m+1}$.
3. S is the set of terminal nodes in $\mathcal{G}_n$

In other words, $D \vdash S$ if and only if there is some derivation $\mathcal{G}$ such that S is the set of terminal nodes in $\mathcal{G}$.

24.5 Soundness

Definition 12 A transformation rule R is *valid* if and only if, if $D \leadsto S$ by R then $D \models\!\!\subset S$.

Rules 1, 2 and 3 are valid. This set can be extended to include rules corresponding to all of Hilbert's axioms.[4]

θ If $D \vdash S$ then $D \models\!\!\subset S$

Proof. Suppose that $D \vdash S$. Then there is a sequence $\Gamma =< \mathcal{G}_0, ..., \mathcal{G}_n >$ that meets the conditions listed in Definition 11. We need to prove $D \models\!\!\subset S$. The proof will be by induction on the length of Γ.

1. Basis case
   The length of Γ is 1.
   Then $S = \{D\}$.
   Therefore $D \models\!\!\subset S$.
2. Inductive step
   Let S be the set of terminal nodes of $\mathcal{G}_m$ and S' be the set of terminal nodes of $\mathcal{G}_{m+1}$. We need to prove that if $D \models\!\!\subset S$ then $D \models\!\!\subset S'$.
   Suppose $D \models\!\!\subset S$ and $\mathcal{M} \models D$.
   Then there is a $\mathcal{N} \supseteq \mathcal{M}$ such that $\mathcal{N} \models E$ for some E in S (by Definition 8).
   By Definition 11, either E is in S' or there is a S'' such that $E \leadsto S''$ and $S'' \subseteq S'$.
   Suppose E is in S'. Then $D \models\!\!\subset S'$.
   Suppose there is a S'' such that $E \leadsto S''$ and $S'' \subseteq S'$.
   Then, by the validity of the rules of transformation, there is a $\mathcal{N}' \supseteq \mathcal{N}$ such that $\mathcal{N}' \models E'$ for some E' in S''.
   Hence, by transitivity of $\supseteq$ there is a $\mathcal{N}' \supseteq \mathcal{M}$ such that $\mathcal{N}' \models E'$ for some E' in S' (since $S'' \subseteq S'$.)
   Therefore, $D \models\!\!\subset S'$. □

References

Barwise, Jon, and John Etchemendy. 1990. Information, Infons, and Inference. In *Situation Theory and Its Applications, I*, ed. Robin Cooper, Kuniaki Mukai, and John Perry. 33–78. CSLI Lecture Notes Number 22. Stanford: CSLI Publications.

Hilbert, David. 1902. *Foundations of Geometry.* La Salle, II: Open Cout Classics (1971).

Luengo, Isabel. 1995. *Diagrams in Geometry.* Doctoral dissertation, Indiana University.

[4]For an example of such a set and the corresponding proofs of validity I refer the reader to my PhD dissertation, Luengo 1995.

25

Belief as a Form of Subjective Information

DANIEL MACK

Abstract

We explore the relationship between notions of belief and of information. We argue that belief can be regarded as a subjective form of information, and derive a formal model based on this conjecture. We show how our model relates to proposals about belief and information in the literature. We argue that it possesses some basic properties identified for belief in those sources, and show that a notion of knowledge naturally arises in our consideration of these properties.

Introduction

This paper is an attempt to model the intuition that beliefs have an "information" content. We work closely with the explanation of this relationship given by Dretske (1981), but we develop a formal set-theoretic model whose properties are more readily open to examination (see, for example Mack (1994a)).

25.1 Existing Work

25.1.1 Dretske's Theory of Belief

Dretske regards beliefs as being properties of cognitive structures within the agent. The theory can be summarised in five steps:

- (Dretske 1981, p. 57) Cognitive structures have a propositional *information content* which is objective, true, and quantifiable in

Supported by EPSRC PhD award number 9131019X.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

a similar fashion to the formal notion of information proposed by Shannon and Weaver (1949).

- (Dretske 1981, p. 173) Knowledge and belief are highly intentional: for example even if two propositions ϕ and ψ are analytically equivalent, belief in one does not entail belief in the other.
- (Dretske 1981, pp. 175-179, 184) Corresponding to each cognitive structure is a unique *semantic content* which is the effective "meaning" carried by the structure and corresponds to the "outermost informational shell" (see below).
- (Dretske 1981, pp. 86-92) *Knowledge* consists of those beliefs which are causally sustained by (true) information; it can be equated with the semantic content of certain cognitive structures.
- (Dretske 1981, pp. 192-193) *Beliefs* are not equivalent to the semantic content of cognitive structures. New cognitive structures can be developed by the agent from previous ones which carry knowledge; the beliefs associated with a new structure are derived from the semantic content of its "parent".

Cognitive structures can only have the semantic content that some proposition is true if their existence is totally constrained by the truth of that proposition and the laws of nature. The notion of "outermost informational shell" requires some explanation: Dretske gives the example of the information that P is a square. If a structure has this as part of its information content, it also has the information that P is a rectangle, and that it is a parallelogram, and that it has four sides (remember, this information is *objective*, independent of any scheme of individuating objects in the world). He says that this subsidiary information is "nested inside" the semantic content of the structure. The "nesting" occurs because given the constraints which exist between what we mean by being a square and what we mean by being a rectangle or parallelogram, the latter propositions follow from the objective truth of the semantic content.

We agree with many of Dretske's insights, and develop them further in our model below. However, we feel that his model of belief becomes unclear where he talks about deriving new cognitive structures from previous ones. Although what he says may correspond with some models of human concept formation, we feel that this part of his theory may in fact be too restrictive of what we can legitimately count as beliefs.

25.1.2 Models of Semantic Information

Two of the core concepts of Dretske's theory are *information* and *semantic content*. In developing a formal model based on his approach, a first step is to examine existing formalisations of these or similar notions.

25.1.2.1 Bar-Hillel and Carnap

Bar-Hillel and Carnap (1964) develop a formal model of *semantic information* which has similar properties to Dretske's notion of information. They work with states of affairs which can be captured by a simple logical language consisting only of *individuals* (which represent objects in the world) and one-place *primitive predicates* which represent distinct properties of the objects. The semantic information of one particular state of the world is defined in terms of the states of the world which can be described using the given individuals and predicates. In order to do this, first they define a *Q-predicator* as a conjunction of all the primitive predicates, where each is either unnegated or negated (not both). For example, if there were two primitive predicates P_1 and P_2, then one possible Q-predicator would be (using lambda abstraction) $\lambda x.\ (P_1 x \wedge \neg P_2 x)$. A *Q-sentence* is a Q-predicator applied to an individual. The set of Q-sentences for a particular individual capture all the states of that individual which can possibly be described in the given language. A *state-description* is a conjunction of Q-sentences, one for each individual; thus the state-descriptions cover all states of the world which can possibly be described. The *semantic information content* of a sentence $\imath$ is defined as the set of state-descriptions in which $\imath$ does *not* hold. Bar-Hillel and Carnap show that this accords with the intuition that for two sentences $\imath$ and $\jmath$, if $\imath \Rightarrow \jmath$ then the semantic information content of $\imath$ is a superset of that of $\jmath$.

Their model of semantic information is *objective*; it depends only on the possible states of the world (which they interpret as those states that can be described), not on any interpreting agent. The semantic information is constructed from propositions, and can be quantified with a measure which has the properties of Shannon's information measure (we have omitted the details of this). As such, it seems a very suitable candidate for a formalisation of Dretske's notion of information. However, they restrict their definition of semantic information to a language containing only one-place predicates. This is because of problems enumerating the set of state-descriptions where predicates of higher arity are allowed (see Carnap (1962)). We believe that where such an enumeration is not required (for example, where quantifying the amount of information is not important), it is justifiable to extend the model to more complex languages.

25.1.2.2 Rosenschein

A different model of information content is proposed by Rosenschein (1985). It is defined using a model of an agent as a deterministic automaton $\langle S, \Sigma, A, \delta, \lambda, s_0 \rangle$, where the parts of interest are:

- S is a set of *agent states*
- Σ is a set of *inputs*

- $\delta : S \times \Sigma \to S$ is a *next-state function*
- $s_0 \in S$ is an *initial agent state*

The agent is set in an environment which can be in one of a very large number of *world states*. The initial world state is represented as ϕ_0, and the state of the world after a sequence $\overline{\sigma}$ of inputs to the agent as $\phi_0/\overline{\sigma}$. For the purposes of this account, world states are assumed to be propositions. In order to find the set $\mathcal{I}_s$ of possible input sequences which lead the agent to be in a given state s, we extend the next-state function δ to a new function $\overline{\delta}$ which operates on an initial agent state and a *sequence* of inputs:[1]

$$\begin{aligned}\overline{\delta}(s, \langle\rangle) &= s \\ \overline{\delta}(s, \overline{\sigma}; \sigma) &= \delta(\overline{\delta}(s, \overline{\sigma}), \sigma)\end{aligned}$$

We can now define $\mathcal{I}_s$ as $\{\overline{\sigma} : \overline{\sigma} \in \Sigma^* \wedge \overline{\delta}(s_0, \overline{\sigma}) = s\}$.

Rosenschein now defines as a pair of related concepts $\mathcal{I}(\psi)$, the set of input sequences after which the proposition ψ is true, and $\mathsf{Knowledge}(\mathcal{I})$, the most that can be known about the world given one of a set of possible input sequences $\mathcal{I}$:[2]

$$\begin{aligned}\mathcal{I}(\psi) &= \{\overline{\sigma} \in \Sigma^* : \phi_0/\overline{\sigma} \Rightarrow \psi\} \\ \mathsf{Knowledge}(\mathcal{I}) &= \bigvee_{\overline{\sigma} \in \mathcal{I}} (\phi_0/\overline{\sigma})\end{aligned}$$

He defines the *information content* of an agent state s as $\mathsf{Knowledge}(\mathcal{I}_s)$.

At first sight, Rosenschein seems to have less justification for his use of the term "information" than Bar-Hillel and Carnap. However, his $\mathcal{I}(\psi)$ is remarkably similar to the set complement of their semantic information content; a feature which we make use of later. His notion of information content is objective and can be defined in terms of propositions, so in this sense is suitable as a formalisation of Dretske's information. In addition, Rosenschein supplies the intuition that information content is not just dependent on the language we choose to describe the world: it is also dependent on real *constraints* concerning how things in the world can change (in this case the next-state function of the agent).

25.1.3 Belief and Semantic Information

All the models we have surveyed stress the objective nature of semantic information, whereas we are intending to model belief, which is something wholly subjective to an agent. Dretske begins to build a bridge between subjective beliefs and objective information by talking about cognitive structures that in some sense inherit their meaning from others; however, he

[1] We use ";" to indicate adding an element to the end of an existing sequence.

[2] We have modified the syntax, but not the semantics of Rosenschein's definitions in order to avoid unnecessary complexity.

provides no details of how or when such structures might be derived. We suggest an alternative approach: the development of a *subjective* notion of information, which relates to belief in the same way as semantic information relates to knowledge. In our model, all cognitive structures will have equal status in terms of evaluating their "information" content. Following a suggestion in the Bar-Hillel/Carnap paper, we shall refer to our new concept as *pragmatic information.*

25.2 A Formal Apparatus For Our Model

25.2.1 States of Affairs

In this section, we construct some different types of "state of affairs" which can be used to represent world and agent states and how they change through time. We shall define an *infon* to correspond intuitively to a simple statement about the world; a *situation* as a set of simple statements which hold simultaneously; and a *course of events* as an association of situations with particular times. These terms are borrowed from the work of Barwise and Perry (1983) and Devlin (1991) on situation theory; however, it is important to note that our *definitions*, though similar, are not identical.[3] As courses of events are manipulated quite frequently in what follows, we also define some basic operations on them.

25.2.1.1 Time and Individuals

Let the (infinite) set of all *time points* be Times, with a total order $\leq$ on its members. We define the *next time point* after t, $\hat{t}$ by:

$$\hat{t} \in \mathsf{Times} \wedge \hat{t} > t \wedge \forall u \in \mathsf{Times}.\ (u \leq t \vee u \geq \hat{t})$$

Let us define a set Atoms of arbitrary but unique identifiers, known as *atomic individuals*, and a superset of Atoms, the *individuals* ($\mathsf{Individuals}$).

25.2.1.2 Infons and Situations

Let an *infon* be a pair of which the first member is an individual known as a *relation* and the second, (the *argument*) is a tuple of individuals. Let an arbitrary set of infons be a *situation* and the set of all possible situations be Sits.

Let each atomic individual α have an associated non-empty set of (non-empty) *atomic situations*, $\mathsf{AtSits}(\alpha)$. The atomic situations can be regarded as descriptions of the possible "states" of the atomic individuals. Let the set of all atomic situations be AtSits.

For each relation r, let there be a unique *negative relation*, written $\overline{r}$, such that $\overline{\overline{r}} = r$. So, if r stands for "likes", then $\overline{r}$ will stand for "does

[3]This borrowing of mathematical structures is unsurprising considering the aim of situation theory is to relate meaning to a notion of information content derived from Dretske

not like". We say a situation is *conflict free* if it contains no instance of an infon and its negation. Let the set of conflict free situations be $\mathsf{FreeSits} = \{s : s \in \mathsf{Sits} \wedge \forall \langle r, a\rangle \in s.\ \langle \overline{r}, a\rangle \notin s\}$.

25.2.1.3 Courses of Events

Let a *course of events* be a partial function from Times to Sits: the domain need not be a continuous set of time points. Let the set of all possible courses of events be CoEs, then $\mathsf{CoEs} = \mathsf{Times} \rightarrow \mathsf{Sits}$. Let the set of *atomic* courses of events be $\mathsf{AtCoEs} = \mathsf{Times} \rightarrow \mathsf{AtSits}$. A course of events is *conflict free* if all the situations in its range are members of FreeSits; let the set of conflict free courses of events be FreeCoEs. It is useful to define a function At which forms a course of events from a situation s by associating s with a time t so that $\mathsf{domain}(\mathsf{At}(t, s)) = \{t\} \wedge \mathsf{At}(t, s)(t) = s$.

We define a relation $\sqsubseteq$ between courses of events c and c' to indicate that c is contained within c':

$$c \sqsubseteq c' \Leftrightarrow \mathsf{domain}(c) \subseteq \mathsf{domain}(c') \wedge \forall t \in \mathsf{domain}(c).\ c(t) \subseteq c'(t)$$

It is straightforward to prove that $\sqsubseteq$ is a partial order and $\langle \mathsf{CoEs}, \sqsubseteq\rangle$ is a lattice. This allows us to define meet and join operations on the lattice, which we shall write as $\sqcap$ and $\sqcup$ respectively:

$$\begin{aligned}
e = c \sqcup d \quad &\Leftrightarrow \quad e \in \mathsf{CoEs} \wedge \mathsf{domain}(e) = \mathsf{domain}(c) \cup \mathsf{domain}(d) \ \wedge \\
&\qquad \forall t \in \mathsf{domain}(c) \cap \mathsf{domain}(d).\ e(t) = c(t) \cup d(t) \ \wedge \\
&\qquad \forall t \in \mathsf{domain}(c) - \mathsf{domain}(d).\ e(t) = c(t) \ \wedge \\
&\qquad \forall t \in \mathsf{domain}(d) - \mathsf{domain}(c).\ e(t) = d(t) \\
e = c \sqcap d \quad &\Leftrightarrow \quad e \in \mathsf{CoEs} \wedge \mathsf{domain}(e) = \mathsf{domain}(c) \cap \mathsf{domain}(d) \ \wedge \\
&\qquad \forall t \in \mathsf{domain}(c) \cap \mathsf{domain}(d).\ e(t) = c(t) \cap d(t)
\end{aligned}$$

25.2.2 Constrained States of Affairs

Rosenschein's model of information described above offers the insight that information is dependent not just on what can be described, but on the constraints which exist between things. We generalise this using a similar notion of constraints to that in situation theory. In this section we define constraints and show how they are used.

25.2.2.1 Constraints

Synchronic constraints specify what things must occur simultaneously. We model them as partial functions from FreeCoEs to FreeSits: this allows us to constrain a situation as true throughout some course of events. We define the set of atomic synchronic constraints, $\mathsf{AtSynch} \subseteq \mathsf{AtCoEs} \rightarrow \mathsf{AtSits}$, and the set of all synchronic constraints, $\mathsf{Synch} \supset \mathsf{AtSynch}$.

Causal constraints connect the state of the world at some time point to what happened in the immediately previous time point. Let a causal constraint be a partial function from FreeSits to FreeSits: if C is a causal constraint, and s a situation in its domain, then $C(s)$ must occur at the next time point after s, so that effects long in the future are mediated via effects at every intermediate time point. We define the set of atomic causal constraints, AtCauses $\subseteq$ AtSits $\rightarrow$ AtSits, and the set of all causal constraints, Causes $\supset$ AtCauses.

We shall assume that all the constraints in Synch and Causes are compatible with one another: combining them will never lead to a contradiction. Formal principles of *monotonicity* and *compatibility* are given in Mack (1994a). We shall regard the non-atomic constraints present in Synch and Causes as constraining the properties of an arbitrary number of individuals which represent abstract objects that are capable of individuation by an agent. We discuss the details of this representation elsewhere (Mack 1994b and 1995).

25.2.2.2 Closure under Constraints

A course of events c_0 is *closed under all constraints* (or simply *closed*) if and only if nothing more can be inferred from its contents all constraints in Synch and Causes. We shall write the closure of c_0 as Closure(c_0): its formal definition and existence proof are give in Mack (1994a). The set of all closed courses of events is defined by ClosCoEs $= \{c : c \in \text{CoEs} \wedge c = \text{Closure}(c)\}$. We also define ClfCoEs, the set of courses of events which are closed and do not make contradictory claims on the world: ClfCoEs = ClosCoEs$\cap$FreeCoEs.

25.2.3 Situated Agents

Finally, we are in a position to define what we mean by an *agent*, the entity which holds pragmatic information and beliefs. We have drawn heavily on Rosenschein's notion of a situated automaton, but our agent states, which we call *cognitive states*, are modelled by situations so that they have internal structure, and are equivalent to the *combination* of current agent state and inputs in the Rosenschein model. The equivalent of the next state function, called the *agent transition function*, returns a partial cognitive state called the *core* of the next cognitive state. The remainder of the next cognitive state consists of the *inputs* to the agent, which are determined by the behaviour of the world. We define what behaviours the world is capable of (its *histories*) in order to show what the inputs to the agent will be.

Let one member of Causes be known as Trans, the agent transition function. Its domain is known as CogStates, the set of cognitive states of the agent, and the members of its range must be subsets of cognitive states, which we shall refer to as *cognitive structures*:

$$\forall s \in \text{CogStates}.\exists s' \in \text{CogStates}.\ \text{Trans}(s) \subseteq s'$$

A *history* is a course of events which extends over all time points, is closed under all constraints, conflict free, and assigns an atomic situation to every atomic individual at every time point and a cognitive state to the agent at at least one time point. We define the set of histories, Hist, by:

$$\begin{aligned}\mathsf{Hist} = \{H : \; & H \in \mathsf{ClfCoEs} \wedge \mathsf{domain}(H) = \mathsf{Times} \wedge \\ & \exists t \in \mathsf{Times}. \exists s \in \mathsf{CogStates}.\; s \subset H(t) \wedge \\ & \forall t \in \mathsf{Times}. [\forall \alpha \in \mathsf{Atoms}. \exists s \in \mathsf{AtSits}(\alpha).\; s \subset H(t)]\}\end{aligned} \tag{1}$$

Let there be one distinguished history Ω called the *actual history*, which represents what happens in the real world in some particular instance of the model.

The cognitive state of the agent in history H at time t is given by:[4]

$$\mathsf{CS}(H,t) = s \Leftrightarrow \forall s' \in (\mathsf{CogStates} \cup \emptyset).\; [s \subseteq s' \subseteq H(t) \Rightarrow s' = s]$$

We refer to the result of applying the transition function as the *core* of the next cognitive state. In history H at time $\hat{t}$, the core and *inputs* of the agent are defined by:

$$\begin{aligned}\mathsf{Core}(H,\hat{t}) &= \mathsf{Trans}(\mathsf{CS}(H,t)) \\ \mathsf{Inputs}(H,\hat{t}) &= \mathsf{Closure}(\mathsf{AllButAgt}(H,\hat{t}))(\hat{t}) \cup \\ &\quad [\mathsf{Closure}(\mathsf{AllButAgt}(H,t))(\hat{t}) \cap \mathsf{CS}(H,\hat{t})] \\ \forall H \in \mathsf{Hist}. \forall t \in \mathsf{Times}.\; & \mathsf{CS}(H,t) = \mathsf{Core}(H,t) \cup \mathsf{Inputs}(H,t)\end{aligned}$$

where we have used the abbreviation $\mathsf{AllButAgt}(H,t') = \mathsf{At}(t',[H(t') - \mathsf{CS}(H,t')])$. We have stated explicitly that the inputs plus the core equals the complete cognitive state.

25.3 Semantic and Pragmatic Information

In this section, we show how a notion of *semantic information* can be formalised within our model, and we compare it with those mentioned above. We show in two stages how to modify it to obtain a notion of subjective, or pragmatic information, and hence belief.

25.3.1 Semantic Information

We define the *semantic information content* of a situation s by:

$$\mathsf{SemInfo}(s) = \{\langle H,t \rangle : s \subseteq H(t)\}$$

$\mathsf{SemInfo}(s)$ has the desired property of being *objective*: although the cognitive states of the agent are used to define the histories, if there were more than one agent, all their cognitive states would have to be included in the histories, so the definition is not dependent on the properties of one particular agent. If we replace the history, time point pairs by Rosenschein's

[4] It is convenient to define the function CS to return the empty set where an agent does not exist at the given time in the given history.

finite input sequences, the situation s by the proposition ψ, and set inclusion by material implication, then this becomes identical to Rosenschein's definition of $\mathcal{I}(\psi)$.

Suppose we now take the complement with respect to $\mathsf{Hist} \times \mathsf{Times}$:

$$\mathsf{SemInfo}'(s) = \{\langle H, t\rangle : s \not\subseteq H(t)\}$$

If we replace the history, time point pairs by Bar-Hillel and Carnap's notion of state descriptions, the situation s by the sentence i, and set inclusion by their notion of logical entailment, then $\mathsf{SemInfo}'(s)$ is equivalent to their definition of semantic information content. However, this analogy requires a little more attention than the previous one, as it is not so exact. The state descriptions of Bar-Hillel and Carnap assume that all combinations of primitive predicates with individuals are logically independent. Our history, time point pairs are indeed full state descriptions of the world, but they are restricted by constraints, thus removing logical independence. In addition, our language of infons involves predicates of higher arity than one, whereas Bar-Hillel and Carnap's model does not deal with these. In Bar-Hillel and Carnap (1964), both of these are recognised as deficiencies of their model. Our model retains theirs as a limiting case, in which a Shannon-like measure can be assigned to it. For these reasons, we regard our model as a natural extension for which we are justified in retaining the term "semantic information content". If s is a cognitive state, then we have a model of the information content (in Dretske's terms) of the agent.

25.3.2 Implicit Pragmatic Information

In this section, we consider how we can modify our model of objective semantic information in order to construct a model of "subjective" information content. We work with the intuition that given certain objects in the world that are objectively real, and certain constraints between them that are objectively real, there are only certain "meanings" which any agent can capture. These meanings are restricted by the capacity of the agent to carry information (which we can imagine as a quantity). However, they are not as restricted as the meanings which are derived from histories (those in the semantic information content), because the agent cannot carry enough information to fully model the behaviour of the world.[5] We capture this notion by reconstructing all those behaviours of things from the world which could cause the agent to be in the cognitive state it actually has at the actual time.

To do this formally, first we define the equivalent of histories, those behaviours of actual objects in the world which yield at least one agent

[5]This is a version of Ashby's law of requisite variety.

cognitive state.[6] We call these the *implicit histories*:

$$\mathsf{ImpHist} = \{h : h \in \mathsf{ClosCoEs} \wedge \mathsf{domain}(h) = \mathsf{Times} \wedge \\ \exists t \in \mathsf{Times}. \exists s \in \mathsf{CogStates}.\ s \subseteq h(t)\}$$

It is easy to see that histories are implicit histories, but not vice versa. We can extend the function $\mathsf{CS}(h, t)$ to apply to all implicit histories h in an obvious way. Unlike the histories, the implicit histories are subjective: they depend only on the cognitive states of the agent under consideration. If there were other agents, each would have its own set of implicit histories, derived from its own set of cognitive states. Using the implicit histories, we can now construct our first attempt at a model of subjective information, known as the *implicit pragmatic information content*:

(2) $\mathsf{IPragInfo}(s) = \{\langle h, t\rangle : h \in \mathsf{ImpHist} \wedge t \in \mathsf{Times} \wedge s = \mathsf{CS}(h, t)\}$

Because its construction mirrors the construction of $\mathsf{SemInfo}(s)$, where we simply replace Hist by $\mathsf{ImpHist}$, it carries over its information-like properties. For example, at least in the limit that only predicates of arity one are allowed, the set complement of $\mathsf{IPragInfo}(s)$ with respect to $\mathsf{ImpHist}$ can be assigned a Shannon-like measure of quantity of information.

Unfortunately, $\mathsf{IPragInfo}(s)$ does not meet our requirements. There can be many implicit histories which are essentially descriptions of the same world behaviour at different levels of abstraction and which still, therefore, result in the same cognitive state at the time of interest. The agent could never even in principle distinguish between these, so they should not be alternative "possible worlds" for belief. We deal with this problem by taking, in each of these cases, only the most abstract implicit history. We shall call the set resulting from this construction the *agent-limited* implicit pragmatic information content:

$$\mathsf{ALIPragInfo}(s) = \{\langle g, t\rangle : \langle g, t\rangle \in \mathsf{IPragInfo}(s) \wedge \forall \langle h, t\rangle \in \mathsf{IPragInfo}(s).\ h \not\sqsubseteq g\}$$

Because of the abstraction present in this construction, the actual history is unlikely to be present in full; instead a simplified version will be present. $\mathsf{ALIPragInfo}(s)$ may use various non-atomic individuals and their properties to summarise various indistinguishable behaviours of the atomic individuals. This begins to address the question of how the model can capture false beliefs. In addition, and once again because of its construction, $\mathsf{ALIPragInfo}(s)$ retains its information-like properties.[7]

[6] Normally this will be as part of a sequence of agent transitions, but we do not assume that the agent exists over all time, so this is not necessarily the case.

[7] For example, a measure of information content could be constructed by considering the set complement of $\mathsf{ALIPragInfo}(s)$ relative to $\mathsf{ImpHist}$.

25.3.3 Explicit Pragmatic Information

It seems that the agent-limited implicit pragmatic information content is a reasonable model for Dretske's notion of information, modified to be *subjective* to an agent: what, then, of his notion of semantic content?

Although we have diverged sharply from Dretske's theory by suggesting the information involved in belief is *pragmatic* rather than semantic information, the argument that beliefs are information held in digital form is still compelling. This can be explained with a reason additional to Dretske's, concerned with how a belief can be expressed in English. Dretske gives the example of a structure carrying the information that some object x is both a square and a rectangle. The digital information content of the structure, and hence the belief of the agent, would then be that x is a square. We argue that it is perfectly plausible for an agent to believe that x is a square without believing that x is also a rectangle, because it may never have acquired the concept of a rectangle. The agent associates some property (its subjective experience of "squareness") with the object x. If it had represented the property of "being rectangular", then it could represent a subset relation between its properties of "squareness" and "being rectangular". In this case, however, its property of "squareness" stands in no relation to any other property. We would describe its belief in English as one that x is a square, despite the agent not being able to relate a square to other similar concepts. This seems to provide strong evidence in addition to Dretske's that beliefs are to be regarded as information in digital form.

Unfortunately, it is not immediately clear how to find the part of the agent-limited implicit pragmatic information content that is carried in digital form. This is because Dretske does not consider the case of disjunctive information. For example, it is not clear whether the digital part of the information "(x is a square AND x is a rectangle) OR (y is a rectangle AND y is a parallelogram)" should be regarded as "x is a square OR y is a rectangle" or as "(x is a square AND x is a rectangle) OR y is a rectangle". In other words, given that the agent will definitely believe in the property of "being rectangular", does this mean that it automatically believes that x is rectangular, even though it believes the more specific information that x is a square?

Earlier versions of our model (Mack 1994a, 1994b) were compatible with the view that the agent does in fact believe that x is rectangular as well as square. This choice was based on the intuition that if a property is *represented* by an agent, it must be represented in all the possible worlds where it holds. However, the author now feels that this analysis is incorrect and that essentially the digital parts of each disjunct (possible world) should be found independently. This is based on the intuition that the particular concept of "being rectangular" which is represented in this case is defined in terms of the individuals with that property: in this case only y. The agent has represented a property belonging to y and one ("squareness") belonging to x, but not any relationship between them. In the very different case where an agent believes that something is both square and rectangular, this argument says that a completely different concept of "being rectangular", that of having four sides and four right angles, is involved.[8]

The advantage of this new approach is that the next step in the development of the model, the formation of the explicit pragmatic information content, becomes much simpler. As suggested above, the aim is to abstract away from those infons which are totally dependent on others, by finding the "smallest" part of $\mathsf{ALIPragInfo}(s)$ that together with the constraints is sufficient to reconstruct the original. This can be thought of as the *digital* content of s.

This procedure is applied to each member of $\mathsf{ALIPragInfo}(s)$ individually. Where there are several candidates for the "smallest" part, we use the property that the closure of the join is equivalent to the join of the closures. All candidates have identical closures, so joining these has no effect. Thus the join of all candidate "smallest" parts is also a suitable "smallest" part, which we shall call an *explicit history.* More formally, the candidate "smallest" parts of a course of events e are given by:

$$\begin{aligned}\mathsf{Mins}(e) \quad = \{m: \quad & m \in \mathsf{CoEs} \wedge \mathsf{Closure}(m) = e \ \wedge \\ & \forall m'.\ m' \sqsubseteq m \Rightarrow \mathsf{Closure}(m') \neq e\}\end{aligned}$$

The the explicit pragmatic information content of an agent in cognitive state s, and the set of explicit histories are given respectively by:

$$\begin{aligned}\mathsf{EPragInfo}(s) = \{\langle \eta, u\rangle : \exists \langle h, u\rangle \in \mathsf{ALIPragInfo}(s). \\ \eta = \textstyle\bigsqcup_{m\,\in\,\mathrm{Mins}(h)} m\} \\ \mathsf{ExpHist} = \{\eta : \exists h \in \mathsf{ImpHist}.\ \eta = \textstyle\bigsqcup_{m\,\in\,\mathrm{Mins}(h)} m\}\end{aligned} \tag{3}$$

Clearly the members of $\mathsf{EPragInfo}(s)$ are explicit histories, and a measure of explicit pragmatic information content can be constructed in relation to the full set of explicit histories. It is easy to show that the explicit pragmatic information content possesses Dretske's *third order* of intentionality (Dretske 1981, p. 173): an infon j may well be constrained to be true whenever the infon i is true, but when $\mathsf{EPragInfo}(s)$ is formed, i can be present without j.

25.3.4 Knowledge and Belief

The explicit pragmatic information content still models the agent's uncertainty: the individual courses of events are those world-descriptions which

[8]This corresponds, for example, to the childhood experience of *learning* that a square was a type of rectangle, and learning the reason why.

are possible for the agent, given its cognitive state. So far, the contribution of this paper has been to explain what the agent's world description can actually consist of. We now assert that an agent's *beliefs* are identical to the world-descriptions that are possible to it. We begin by transferring Rosenschein's definition of knowledge to the model:

Definition 1 An agent *knows* an infon $\imath$ at time t when in cognitive state s if and only if $\imath$ occurs at time t in every member of $\mathsf{EPragInfo}(s)$.

We now define what are essentially ordinary and uncertain beliefs:

Definition 2 An agent *believes* an infon $\imath$ at time t when in cognitive state s if and only if $\imath$ occurs at time t in at least one member of $\mathsf{EPragInfo}(s)$, and $\bar{\imath}$ does not occur at t in any member of $\mathsf{EPragInfo}(s)$

Definition 3 An agent *doubts* an infon $\imath$ at time t when in cognitive state s if and only if $\imath$ occurs at time t in at least one member of $\mathsf{EPragInfo}(s)$, and $\bar{\imath}$ occurs at t in some other member of $\mathsf{EPragInfo}(s)$

25.4 Basic Properties of the Model

A belief by an agent in an infon $\imath$ at time t is *true* if the infon $\imath$ occurs at time t in the actual history. The belief in $\imath$ is *false* if the infon $\bar{\imath}$ occurs at t in the actual history. Otherwise, the belief is *indeterminate.*

Theorem 1 *All knowledge is true.*

Proof. To prove this, suppose that the agent in cognitive state s knows i at time t.

$$\forall \langle h, u\rangle \in \mathsf{EPragInfo}(s).\ \imath \in h(t)$$

From this it follows (from equations 3 and 25.3.2) that:

$$\forall \langle h, u\rangle \in \mathsf{ALIPragInfo}(s).\ \imath \in h(t)$$
$$\forall \langle h, u\rangle \in \mathsf{IPragInfo}(s).\ \imath \in h(t)$$

From equations 2 and 1, it follows that every member of Hist is also a member of $\mathsf{ImpHist}$, so the set of histories compatible with the cognitive state s must be a subset of the set of *implicit* histories compatible with s. Thus any property which holds for all the implicit histories in the implicit pragmatic information content also holds for these histories. As the actual history is one of the members of the implicit pragmatic information content, $\imath$ occurs at t in the actual history, and the theorem is proved. $\square$

Theorem 2 *If the agent knows an infon, then it believes it.*

Proof. The proof is straightforward from definitions 2 and 1. $\square$

Theorem 3 *The agent can have false beliefs.*

Proof. Suppose that in the actual history Ω, the infon i occurs at time t: that is $i \in \Omega(t)$. Let the current time be e, and the agent's cognitive state be s, then $\langle \Omega, e \rangle$ must be a member of $\mathsf{IPragInfo}(s)$ (since histories are implicit histories).

Consider two cases. For the first, suppose that there is a member ω of $\mathsf{IPragInfo}(s)$ such that $\omega \sqsubseteq \Omega$ and $i \notin \omega(t)$; suppose also that there is no member of $\mathsf{IPragInfo}(s)$ apart from Ω which has i occur at time t. Then from equation 25.3.2, $\Omega \notin \mathsf{ALIPragInfo}(s)$ because ω is standing in for it. Thus there is no member of $\mathsf{ALIPragInfo}(s)$ where i occurs at t. Note that nothing has been assumed to prevent there being a member of $\mathsf{ALIPragInfo}(s)$ where $\bar{\imath}$ occurs at t. If this is now assumed, then provided $\bar{\imath}$ occurs at t in the derived explicit history, then $\mathsf{EPragInfo}(s)$ has a member where $\bar{\imath}$ occurs at t, but no member where i occurs at t, and the result is proved.

For the second case, suppose that there is no ω with $\omega \sqsubseteq \Omega$ and $\omega \in \mathsf{IPragInfo}(s)$. Then $\Omega \in \mathsf{ALIPragInfo}(s)$. Suppose that the explicit history derived from Ω is Ω' and that $i \notin \Omega'(t)$. Suppose also that there is a member of $\mathsf{ALIPragInfo}(s)$ where $\bar{\imath}$ occurs at t, and that $\bar{\imath}$ still occurs at t in the explicit history derived from this. Then $\mathsf{EPragInfo}(s)$ has a member where $\bar{\imath}$ occurs at t but no member where i occurs at t. Once again, the result is proved. □

The two cases outlined offer insight into how false beliefs arise. In the first case, the agent is incapable of capturing sufficient information about the real world. In the second case, the information is not available *in digital form*, which means the agent cannot access or manipulate it.

These theorems demonstrate formally why we have associated knowledge and belief with the chosen constructs. Knowledge is identified with formally certain belief. Ordinary beliefs are not necessarily formally certain, but their negation must be formally absent.

In the following example, we show how this accords with everyday belief reports which do not allude to either certainty or uncertainty in belief. When Lorraine was a child she believed in Father Christmas (Santa Claus). There was no sense in which she disbelieved in him, but it is still difficult to say what she would have replied if asked whether she was *certain* that Father Christmas existed. She might simply have said there was no reason *not* to believe in him, but might also have realised the possibility that her parents were lying when they told her that Father Christmas distributed the presents on Christmas Eve (especially if this possibility was suggested to her). In both these cases, she would have become less certain in her belief that Father Christmas existed. It could be argued that she was never certain, because she always had the capability to imagine ways in which he could not exist that were compatible with the evidence of her experience. If, under those circumstances, she had said she was certain, she would merely have meant there was evidence in favour but none against the existence of Father Christmas. Thus it appears that in accordance with the model, beliefs we think of, loosely, as "certain" are not in fact formally certain. It also accords with our identification of formal certainty with knowledge: Lorraine would not have said she *knew* that Father Christmas existed.

25.5 Conclusion

We have constructed a mathematical entity, the explicit pragmatic information content of an agent, which has both information-like and belief-like properties. It satisfies our intuition that beliefs are a form of information, showing how this information differs from the previously developed notion of semantic information. Beliefs which are formally certain behave like knowledge. In Mack (1995), we show that our model has some of the other expected properties of belief.

Acknowledgements

The author would like to acknowledge the benefit of comments from Jim Doran, and from participants at the Conference on Information-Oriented Approaches to Language, Logic and Computation, held at St. Mary's College, California in June 1994.

References

Bar-Hillel, Yehoshua, and Rudolf Carnap. 1964. An outline of a theory of semantic information. In *Language and Information*, ed. Yehoshua Bar-Hillel. Addison Wesley.

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes.* MIT Press/ Bradford Books.

Carnap, Rudolf. 1962. *Logical Foundations of Probability.* Chicago University Press. Second edition.

Devlin, Keith. 1991. *Logic and Information.* Cambridge University Press.

Dretske, Fred I. 1981. *Knowledge and the Flow of Information.* Basil Blackwell.

Mack, Daniel. 1994a. Belief as Pragmatic Information: a New Formal Model. In *Proceedings of the 1993 Workshop on Cooperating Knowledge Based Systems*, ed. S. Misbah Deen. Keele University Press.

Mack, Daniel. 1994b. A New Formal Model of Belief. In *Proceedings of the Eleventh European Conference on AI (ECAI-94).* John Wiley.

Mack, Daniel. 1995. *An Information-Oriented Model of Agency.* Doctoral dissertation, University of Essex. Forthcoming. All details provisional.

Rosenschein, Stanley J. 1985. Formal Theories of Knowledge in AI and Robotics. *New Generation Computing* 3:345–357.

Shannon, Claude E., and Warren Weaver. 1949. *The Mathematical Theory of Communication.* University of Illinois Press.

26

Diagram Contents and Representational Granularity

KENNETH MANDERS

26.1 Is Information Representation-sensitive?

The notion of representational content is a difficult one, for many reasons Many of its problems are inherited by contemporary notions of information Notably, one encounters a systematic ambiguity among the standings of an occurrence in different ensembles of alternatives or in differently conceived causal settings, corresponding to different ways in which the occurrence could be understood to convey content The problems are compounded by the non-trivial cross-ensemble identifications of occurrence and content that we routinely make when employing representations [1]

For example, consider the cartoon "The flow of information" in *Situations and Attitudes* (Barwise and Perry 1983, p 17), a sequence from dog-with-broken-leg through veterinary consideration to Jonny's knowing the diagnosis The text asserts " the doctor's utterance provides Jonny with *the same information* ", the same information (in some unproblematic sense) is allegedly present, and represented, at each stage of this chain

But in nature dogs do not come labelled with unique medical diagnoses A dog presents a most variegated content A good veterinarian sees much of diagnostic import that I would miss, a breeder much else The designer of

The author gratefully acknowledges support for this research from NSF grants DIR-90 23955 and SBER94 12895, and a Howard Foundation fellowship for AY 94/95

[1] Concepts of action and responsibility are subject to corresponding difficulties of denomination, which have been more widely noticed Thus "one and the same act" may be construed (for example) in terms of its causal character, in terms of its cause/effect relations as the agent represents them to himself (his intentions), or relative to moral or social requirements Judgements of responsibility can come out quite differently depending on the mode of evaluation

Logic Language and Computation
Jerry Seligmann and Dag Westerstahl eds

the cartoon evidently sensed a problem of this sort, for the initial dog-image was labelled, "Jackie has a broken leg." In this way, the "information" in the dog-image is *linguistically denominated:* a content is assigned as individuated by the standards of linguistic representation. There is a plausible sameness of information or content at each stage, as opposed to some looser content-relatedness, only if we impose the representation of the final stage on the earlier stages. Compared to the natural occurrence allegedly with the same content ("natural meaning"), an artist's sketch locates the occurrence in a very special ensemble of alternatives (sketches), and the artist exercises considerable "conceptual" control over its content in doing so.

For the term 'information', such systematic ambiguity is abetted by the communication engineer's quantitative conception of information (Shannon entropy and channel capacity), which succeeded marvellously in separating engineering concerns from those of the user. The semblance of representation-neutrality of information so conceived arises from the combination of bona fide representation-neutrality of the statistical measure with an artifact of theoretical casting.

Genuine application of the Shannon theory requires individuating the events of interest, to which the statistical measures are applied, at each stage of the communication channel. On the user side, for example, telephone carriers put major research into picking out components of a voice signal most important to understanding transmitted speech, so that others could be filtered out; events relevant to determining channel capacity requirements are individuated after this reduction rather than before. On the channel engineer's side, one has to specify the correspondence between physical receiver states and output states before the entropy calculations relevant to effective channel capacity can be made.

Such specialized tasks figure little in an outsider's conception of the Shannon theory, leaving the mis-impression that individuation of events in statistical ensembles at various stages of the channel is intrinsically given, and that information in the Shannon sense can be unproblematically identified between various stages of a channel.

This systematic ambiguity among the standings of an occurrence in different ensembles of alternatives can give the impression that individuation of contents, and hence *content,* is representation-neutral. Such presumed representation-neutrality of content is often in play in philosophy of language, including Frege's conception of *thoughts* as contents shared among all mankind, and a widespread use of the term 'proposition' to denote translation-invariant linguistic content, which has implicitly served as a philosophical standard for all representational content.

One basic thrust of situation semantics has been to provide explicit machinery bridging the gap between linguistic surface form and content:

situations, semantic evaluation of sentences with respect to situations. Admittedly, the intuitive notion of a situation does not specify any particular representation in which to describe situations. In the presentation of the machinery, however, situations have been almost exclusively linguistically denominated (as, for example, in *Situations and Attitudes*). For purposes of computational linguistics, this is presumably just as well, but it is at least in principle possible that efficiency of some human language processing (not to speak of differences relevant to philosophical conceptions) requires processes without counterpart in ordinary language representation.

While we do not address language processing issues, we want to show that contents cannot always be understood as unproblematically representation-neutral, by considering examples of geometrical representation. It is not clear at the outset to what extent all diagrammatic representational systems have the same general features. For this reason, we focus on one: diagrams in traditional geometry. We first show how the individuation of diagram contents in traditional geometrical practice is intimately bound up with the workings of that practice, and differs from the individuation of corresponding contents in other geometrical representations. We then attempt to get some theoretical grip on this behavior and its consequences for the notion of representational content.

26.2 Diagrams in Traditional Geometry

As does Leibniz, I treat diagrams as *textual* components of a traditional geometrical text or argument, rather than semantic counterparts.

> "[Geometrical] figures must also be regarded as characters, for the circle described on paper is not a true circle and need not be; it is enough that we take it for a circle."
>
> (Leibniz 1969, p. 84)

Aside from the long-standing philosophical difficulties that arise from the assumption that the discursive text is somehow true of the diagram, or a "perfect version" of it, the assumption seems incompatible with the use of diagrams in proof by contradiction. This simple-minded objection has nothing to do specifically with geometry: proofs by contradiction never admit of semantics in which each entry in the proof sequence is true (in any sense that entails their joint compatibility).

I therefore take it that traditional geometric discourse (or text) has a verbal part, the *discursive text,* and a graphical part, *diagrams.* The discursive text consists of a reason-giving ordered progression of assertions *(attributions)* and textual cross-references to features of the diagram. A step in the geometric sequence is licensed by attributions either in the discursive text or to the diagram, or both. A step consists in an attribution in the discursive text, or a construction in the diagram, or both: "With center A and distance AB let the circle BCD be described" –Euclid I 1 (Euclid 1956).

Geometric attribution standards —the balance between diagram-based and discursive text-based attributions— are rarely discussed in the extant ancient texts. The attribution standards, on which we anyhow cannot expect full unanimity over the centuries of Greek mathematical practice[2], nonetheless become intelligible when we consider the demands of coherence.

Mathematical practices (as many other intellectual practices) aim to secure unqualified assent of participants. Among other things, this requires especially strong performance on *coherence:* if a practice licenses a type of response to representations, it must provide means by which differences among participants, in actual response to those physical representations which they might reasonably produce, can be resolved. Otherwise, disarray results, and inhibits unqualified assent. It does not go without saying that a practice has adequate representational facilities to avoid disarray. The division of responsability between diagrammatic and discursive representations in traditional geometry is a response to this challenge. Subsequent geometrical representations are alternative responses.

26.2.1 Exact vs. Non-exact Attributes

Specifically, geometric attribution standards may be understood by distinguishing *exact* and *non-exact* attributes. Exact attributes are those which we, since Descartes' time, could express by equations. In traditional geoemetry, many of them were expressed or defined by equalities or proportionalities. Prominent examples include equality of lines (segments), angles, or other magnitudes, congruence of triangles or other figures, proportionality of lines; that an angle is *right* (not, that it is an angle), 4-point con-cyclicity relationships; the geometric character of lines or curves: that a line is *straight,* that a curve is a *circle,* an *ellipse*; that lines are *parallel;* that three lines or curves intersect in a point (rather than intersecting pairwise in three distinct points); that a line is tangent to a curve (rather than intersecting it in two or more points close together).

Exact attributes are unstable under perturbation of a diagram, well beyond our control in drawing diagrams and judging them by sight. Thus, diagramming practice by itself provides inadequate facilities to resolve discrepancies in response by participants as they draw and judge diagrams.

[2]The most informative source is perhaps Proclus' *Commentary on the First Book of Euclid* (Proclus 1970). For discussion, see especially Heath (Euclid 1956, vol. 1) and Mueller (1981)). Not even all details of ancient diagrams themselves can be ascertained from the late copies in which we possess the texts (Tropfke 1922–1924, vol. 4, p. 14–19): lettering of points is clearly ancient; marks to indicate equal line segments or angles seem to be a 19th-century innovation.

It is therefore unsurprising that exact attributions is licensed only by prior attributions in the discursive text; they may never be "read off" from the diagram. We observe this dependence on attributions in the discursive text throughout ancient geometrical texts. That the curves introduced in the diagram in the course of the proof of Euclid I 1 are circles, for example, is licensed by Postulate 3 in the discursive text, and is recorded in the discursive text to license later exact attributions: equality of radii (by Definition 15, again in the discursive text).

Exact attributions license a variety of extremely powerful inferences, which are pervasive and central in traditional geometrical discourse. Besides substitution inferences:

$$a \simeq b\,,\ \Phi(a) \quad \Rightarrow \quad \Phi(b)$$

which notably include transitivity inferences, we find equalities licensed by definitions (circle has equal radii), and congruence attributions, and licensing congruence attributions. We frequently find proportionality attributions licensed by the various rules for manipulating proportionalities, similarity of figures, and 4-point con-cyclicity (by Euclid III 35–37). Without such inferences (perhaps in some other form), spatial reasoning is handicapped beyond recognition.

Non-exact attributes express recognition of regions (and their lower-dimensional counterparts, segments and points) and their contiguities in the diagram. We might say they express the topology of the diagram. Prominent examples include non-empty delimited planar regions: triangles, squares,... (but not that sides are straight); circles (but not that they are circular rather than elliptical or just irregular); angles (must be less than two right angles, to delimit a region). In lower dimension, non-empty segments (but not the character of the curve of which they are segments); points as two-place (but not three-place or tangent) intersections of curves; *non-parallel* lines; non-tangencies. Further, contiguity and inclusion relations among these: point lies within region; point lies within segment; side *opposed* to vertex; line *divides region* into two parts; triangle *lies within* triangle; *alternate* angles, and so on.

Non-exact attributions are licensed by the diagram directly; for example, the attribution of an intersection point of the two circles in Euclid I 1. This poses no immediate threat for coherence, because non-exact attributes are "locally invariant" under variation of the diagram: they are shared by a range of perturbed diagrams. The allowable range of variation differs. Some attributions, such as that of an intersection point in I 1 or in Pasch's axiom, arise from fundamental topological properties of the Real plane and are extremely stable. On the other hand, in complicated configurations the topology of the diagram and the ordering of salient points on a line can be quite sensitive to perturbation of the diagram.

In any case, the local invariance of non-exact attributes makes it feasible for geometrical practice to attribute them from diagrams, without undue risk of disarray. Diagram-based moves were open to certain challenges already in antiquity, as we see from *objections* recorded by Proclus and others; but the moves were defended and retained their license. By the end of the 19th century, one does indeed want to consider alternative geometries such as those with restricted coordinate domains, and traditional diagrams are indeed unable to provide representational support for reasoning of this type. But this challenge did not arise in traditional geometry. A much more acute threat to coherence in traditional geometry is sensitivity of non-exact relations to variation in complicated configurations. We will return to this.

The program of locating "implicit assumptions" has amply shown that genuinely diagram-based moves occur throughout traditional geometrical argument (Mueller 1981, p. 5, and throughout). With obvious exceptions, almost every step in traditional geometrical argument finds its license partly in the arrangement of the diagram.[3] From a modern foundational point of view (Hilbert 1959, Tarski 1959), it is clear that continuity considerations, in close conjunction with exact relations, contribute much of the existential import to geometry. But the very objects of traditional geometry also arise in the diagram: we enter a line in a diagram, and presto, new triangles pop up to license further moves. Though it has proved remarkably difficult to explain why this is so important, without such inferences, organized rather as they are through the use of diagrams, spatial reasoning is handicapped beyond recognition.

26.2.2 Case Distinctions

Traditional geometrical reasoning distinguishes many cases. Poncelet describes the situation perceptively, if with discontent:

> "The diagram is drawn, and never lost from view. One always reasons about real magnitudes; every conclusion must be pictured... One stops as soon as [in varying the diagram] objects cease to have definite, absolute, physical existence. Rigor is even taken to the point of not accepting conclusions of reasoning on one general arrangement of a diagram, for another equally general and perfectly analogous arrangement. This restrained geometry forces one to go through the whole sequence of elementary reasoning all over, as soon as a line or

[3]Notable exceptions: proportionality manipulation; substitutions with premiss $\Phi(a)$ exact or in any case previously noted in the current discursive text; invocation of previous results, axioms or postulates, the conditions for which have been previously noted.

a point has passed from the right to the left of another, etc "
Poncelet 1822, 1865, p xii–xiii

For example, *De Sectione Ratione* of Apollonius (1987) concerns the problem

> Given two lines, a point on each (Q, Q'), and a point P not on either line, to locate a line through P which (with the given points) cuts off from the given lines segments QR, $Q'R'$ which are in a given ratio

Starting with the major division as to whether the two given lines intersect, Apollonius treats some 87 (!) cases, with occasional further subdivisions Although such complete treatments appear to be rare (Euclid, for example, tends to leave many cases to the reader), Apollonius' careful treatment appears to reflect the highest standard of geometrical reasoning as far as rigor (disarray avoidance) is concerned The standards of traditional diagram-based geometrical reasoning force case distinctions "diagrams individuate claims and proofs"! This we must see

Typical steps in a geometrical proof (at least, many steps in many proofs) respond to the appearance of the diagram, the topological arrangement of its parts As one would expect from a practice that engages its representations in that way, topologically distinct diagrams are treated separately, and topologically indistinct diagrams are treated via a single representative In this way, the representational strategy of the practice succeeds in reducing and making more manageable the diversity of spatial forms (only partly transferring it to the discursive text) Case distinctions arise from this situation in two basic ways

First, there is no general way to predict, from the appearance of the diagram (say, the conjunction of all diagram-based attributions that it licenses without adding auxiliary lines), its appearance after constructions are applied in the course of an argument for this data simply does not determine what topology arises from construction of additional elements Only more detailed metrical data —we can now say, algebraic inequalities— can determine this, and this data is unavailable in diagrams as treated by the practice (namely, not in general attributable, and, as we have seen, for good reason) As a consequence, topologically similar initial diagrams may become dissimilar through constructions in the course of a proof When this occurs, the proof must deal separately with the dissimilar diagrams that have arisen [4]

[4] A liability of traditional practice is that it provides no explicitly discussed systematic representational support for the detection of such alternatives because only one of the initially similar diagrams need be presented until the construction is carried out It seems to have met this liability by providing roles in addition to that

Second, representations other than traditional diagrams (for example, already the traditional discursive text) may allow one to give a single (non-disjunctive) expression to conditions that correspond to a variety of topologically distinct diagrams. Because geometrical argument involves the diagram, proof of such statements requires treating the alternatives separately. In this situation, the alternatives tend to appear either as separate propositions or as separately argued subcases.

For example, when two lines through a point O meet a circle, say, one in A and B and the other in C and D, then the rectangles (products) OA by OB and OC by OD are equal (*cf.* Euclid III 35–37, and Heath's summary of discussion on related subcases and claims, Euclid 1956, vol. 2, p. 71–77). Of the many distinct diagrams which can arise, Euclid III 35 concerns specifically the one where O lies inside the circle, III 36 the case in which O is outside but one given line is tangent to the circle. This requires separate discursive formulation as a proposition because the traditional discursive text cannot be read to include the square on a single segment OP under the product OA by OB. Ultimately, this restriction on discursive representation seems again forced by topological distinction behavior in reading diagrams, given the need for a workable cross-reference system between discursive text and diagram: OP cannot at all be visually located in a diagram showing distinct OA and OB, and in the opposite case one would at least have to allow double-labelling.

Cartesian geometrical representation by equations, and others inspired by it, such as Poncelet's (discussed below), lead us to recognize many case distinctions in traditional geometry in this (second) way: results seem analogous to us, which are treated separately in traditional geometry. Often, ancient texts already show, at least by their expository arrangement, that analogies are sensed among such distinctly treated questions; often, they do not. For example, algebraic representation often allows one to include limiting or degenerate cases (tangencies or coincidence above) in the treatment of the "generic" case. The importance of continuity in this sense was stressed by Leibniz and Poncelet. Traditional diagrammatic representation requires that limiting or degenerate cases be handled separately; and when limiting cases are simpler to treat, it may then be desirable to prove the generic case via limiting cases. Some have held that this is Euclid's strategy in treating the tangency case for the analog of III 35 where O is outside the circle; if so, this forces a disanalogy between the demonstrations of the O-inside and -outside propositions, because when O is inside the circle, it has no (real, traditionally diagrammable) tangent through O.

---

of the mathematical expositor, allowing moves of proposing *case* and *objection* — see Proclus 1970, p. 212, and throughout the subsequent commentary, (Euclid 1956, vol. 1, p. 134–35); and Heath (Euclid 1956, vol. 1, p. 134-35.)

Thus, traditional practice with diagrams controls the individuation of claim and proof; in contrast with Cartesian and 19th-century projective geometrical practices, which exploit different representation to individuate geometrical claims and proofs less finely.

26.2.3 Uniformity Enhanced: Poncelet's Projective Geometry

The Cartesian style of reducing geometrical conditions to algebraic equations turns out to eliminate non-exact features: many diagrams with different topologies lead to the same equation. The resulting strikingly uniform style of geometrical argument calls for a detailed analysis by itself. Instead, we here consider aspects of the influential 19th-century geometrical reaction by Poncelet, who, by developing a new practice of building and reading diagrams, aims to enjoy the unificatory advantages of Cartesian geometry while re-gaining geometrical motivation lost in Cartesian procedure.

Based on extensive deliberations on the geometric correlate of sign changes in algebra culminating in the work of Carnot 1803, Poncelet handles diagrams that are related by direction reversals by suitable changes of sign in additive and multiplicative relationships among line segments. So read, Euclid's original proof of III 35 already covers the O-outside-circle case.

After detailed algebraic preliminary investigations (1982-64, vol. 1), Poncelet found also a way to make diagrams involving conic sections behave like complex-coordinate analytic geometry, thus eliminating traditional geometrical case distinctions in the theory of conics. The trick, intimately related to III 35–37, is to extend each conic by a tangent "supplementary conic" of opposing type —same diameter, same parameter, opposite excentricity (Poncelet 1822, Section Première, Chapitre II). For example, two circles may fail to intersect in traditional geometry; in that case, the extended circles continue to have two "ideal" intersection points (marked in Figure 1). Poncelet's new diagramming practice allows indifference to Euclidean "reality" of intersections of conics, which traditionally requires different diagrams.

The tangency case III 36 also obtains an analogue when O is inside the circle. Extend the circle by its supplementary orthogonal hyperbola with respect to the diameter through O. The hyperbola has two tangents through O, which Poncelet regards as "ideal" tangents to the circle through O. (See Poncelet 1822, figure 6 and Figure 2). The length of these ideal tangents OP' must be understood in a non-Euclidean way, however; it turns out to be representable by the lines OP. Once this is granted, this ideal tangency case is (surprisingly) covered by Euclid's demonstration of the ordinary tangency case III 36 with appropriate sign modifications. The

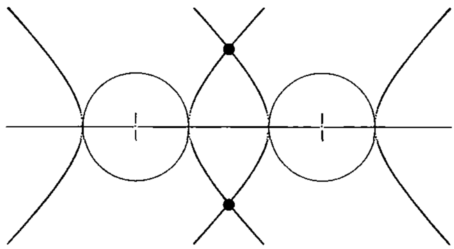


FIGURE 1 Supplementary-conic unification: ideally intersecting circles.

range of statement and proof is thus extended to cover both "real" and "ideal" tangency cases, with surprising uniformity.

26.3 Representational Granularity

26.3.1 Re-individuation and Content

We now encounter a remarkable phenomenon. When Cartesian practice subsumes several distinct traditional geometrical problems under a single equation and subsequent algebraic treatment, it comes to regard those as *the same* problem, one problem. When Poncelet can give a unified statement and treatment for inside-circle and outside-circle versions of Euclid III 35–37, he can say he has made it possible for us to see that those are all *the same* geometrical problem. Diagrams individuate geometrical problems; cartesian algebraic representation re-individuates them (those same problems); so does Poncelet's modified diagrammatic representation.

When a practice treats traditional geometrical problems, rather than starting its own game, representational usage in that practice imposes an individuation on geometrical problems, with approximately the same rights as traditional geometrical diagram representation. Recall how it went: the practice responds to X in its representations (topological arrangement of

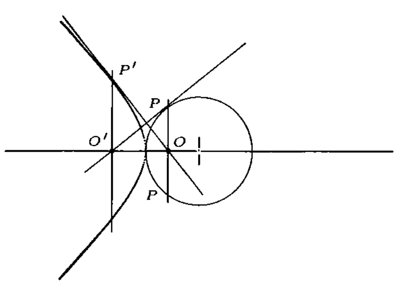


FIGURE 2 Supplementary-conic unification for Euclid III 36.

the diagram); consequently, it treats X-distinct representations (topologically distinct diagrams) separately, and X-indistinct representations (topologically indistinct diagrams) through a single representative.

That practices should individuate contents in this way is unremarkable. Our "remarkable phenomenon" arises because the practices are *linked:* they come with a content correlation practice in virtue of which they count as treating common geometrical problems, which we might say are being re-represented in the alternative practices. Thereby, their claims to control problem and content individuation come into competition. That content individuations so related should disagree is philosophically surprising because in the philosophy of language we have come to expect contents, properly understood (propositions, say), to correspond one-to-one across genuine content-correlations between practices (correct translation, for example). This may be approximately correct for translations among natural languages; but evidently not for re-representation correspondences in mathematics.[5]

Should we accept the claims to common content that arise from our geometrical correspondences? Transitions from traditional geometric practice to Cartesian or Poncelet-style geometry seem anything but revolutionary in the Kuhnian sense: the results of traditional practice are subsumed, not rejected; individual geometric results obtainable by the new method are provable in traditional practice (though not provable in one piece, in the cases at hand). The relationship of shared content thus seems much closer here than between, say, classical and special-relativistic mechanics. Denying commonality of content in the geometrical cases will cost us a gratuitous and counter-intuitive loss of intelligibility of our intellectual world. The alternative practices lose their otherwise obvious geometrical motivation if we do so.

[5]There may be shades of Kripke's puzzle here (Kripke 1979). We take a participant in a practice —as the case may be, traditional geometry; or Pierre's thought— a practice that distinguishes two contents —O-inside-circle segment product theorem and O-outside-circle segment product theorem; or Pierre's Londres/London. We judge the agent by the content individuation standards of a different practice, which can also claim those contents but judges them one —Poncelet's practice, or Kripke's which treats 'Londres' and 'London' as giving rise (via translation) to the same propositions. When we catch the agent taking a different stand on two such contents —our geometer asserting one theorem as just proved, the other as still needing proof; Pierre asserting 'Londres est jolie' and denying 'London is pretty'— the situation looks strangely incongruous. Strangely so, to the extent that the agent's content-distinguishing stance qualifies as arising from a rational intellectual practice: someone might try to duck the puzzle in Pierre's case by regarding his stance as *merely* psychologically coherent; or at least find the story too thread-bare to support decisive judgement on attribution of rational intellectual practice. There ought not to be that problem in the case of traditional diagram-based geometry.

26.3.2 Representational Types and Representational Granularity

The apparent unexpected re-individuation behavior of content under re-representation calls for more analysis of how we use representations.

Practices pursue their aims by *engaging* their representations, such as diagrams, discursive text, or algebraic displays. Engagement consists in generation and acceptance of a succession of physical representations, subject to the standards of the practice (proof steps). In traditional geometrical practices, Euclid's first postulate, taken in conjunction with a diagram to which two distinct points may be attributed, licenses introduction into the diagram of a "line" curve including those two points, as well as subsequent attribution "straight line" to that item in the discursive proof text. Mathematical practices tend to use one or more kinds of representation each, just as traditional diagram-based geometries use diagrams and discursive text jointly. We call such a kind of representation, as governed by the standards of a practice, a *representational type.*

A representational type depends on standards of a practice as well as on the physical characteristics of its representations. For example, every geometrical diagram, as a physical object, inescapably has a determinate size (and further geometrically irrelevant physicality). Traditional geometrical practices require one to engage (produce, respond to) diagrams taking differences in their size to be irreleveant. Thus it can use a single diagram and argument to handle different sizes.

But traditional practice licenses no response insensitive to whether the point is inside or outside, the way it expects a response insensitive to diagram size. For this reason, it must handle separately the cases where the point is inside or outside. In terms of content, we would say that it does not make available a content "point and circle" in diagram and discursive text separately from a content of the kind "outside circle," "inside circle," or "on circle." The contents "point outside circle", "point inside circle", or "point on circle", the configurations are just different; the practice provides no response directly to a shared content "point and circle." Even when it looks to us as if we can express contents separately in the traditional discursive text (did we not just do so: "point and circle"?), pertinent inferences are, within traditional practice, available only via engagement with the whole diagram (an attribution or construction move). This example illustrates a general tendency. The representational type of traditional diagrams (notably, geometrical standards for engagement) often provides for a response only to a relatively rich configuration, compared to what is possible in Cartesian or projective geometry.

Representational types, such as traditional or Poncelet geometrical diagrams, discursive text, and systems of equations, differ in how parts of contents that they share can engaged separately. Such differences we call grain or *granularity* differences between the representational types. In a Poncelet diagram, we can still judge that a point is inside a circle if we need to; but in Poncelet practice, we can engage the diagram indifferently to this and uniformly argue the two-product property considering only a single diagram, where in traditional geometry we cannot. Thus Poncelet-style geometry has a separately engagable content "point and circle" where traditional geometry does not. Poncelet practice unifies the inside-circle and outside-circle versions of III 35 because these results turn only on a shared content which it enables us to engage in argument. Because of the different representational grain of its diagrams, Euclidean practice can only engage that shared content via distinct more inclusive contents.

26.3.3 Partial Contents

Our discussion turns on a part-whole relationship among contents, that is taken as primitive here. This relationship already applies within a single representational type or practice.

In diagram-based traditional geometry, the notion of a sub-configuration already plays a crucial role. To apply a previous result in a geometrical proof, we must pick out the configuration it concerns as a partial content of ("within") the configuration presently considered. Previous results may be applied rather than re-proven for the application because auxilliary constructions in prior proofs are taken not to interact with the diagram of an application. Were they instead carried out in that richer diagram, their interaction with its additional elements would give rise to additional case distinctions. In this way, diagrams would quickly become unmanageably complex, and require so many separate proof cases, characterized by non-exact diagram-based relations which would then be so sensitive to variation of diagram, that traditional geometry would collapse into disarray a few propositions into Euclid. Those striking "long chains of reasoning" depend critically on the stability of partial contents taken separately from a larger configuration, to keep sensitivity to variation of diagram in check.

Nor is this all. Compared to an imaginary proto-geometrical mode of extraction of geometrical truths simply by deliberating on a diagram, sequential geometrical argument itself plays fundamental intellectual roles: justification of course, but also *reason-giving* in the broader sense of making connections of reason for us to grasp. But there are no connections without connectables: separate geometrical claims and attribution statements. These separate geometrical claims and attribution statements are themselves partial contents relative to diagrams.

As arguments get long, however, they become less accessible to our grasp. Understanding is promoted by a further layer of ordering into a dependent sequence of contents, propositions. Propositions can function as free-standing graspable contents, whereas attribution statements (like $AB = AC$) do not: their role in the practice depends on an anaphoric cross-reference relationship to a diagram. It is thus a crucial feature of the representational type of discursive text that it provides free-standing partial contents that can be ordered in this way.

In Cartesian geometry, inclusions among partial contents also play important roles. Book III of the *Géométrie* (Descartes 1954, p. 383–88) gives a striking example. Having obtained from a geometrical problem a quartic equation with algebraic terms formed from the known quantities of the problem as coefficients, Descartes first invokes a theory of factorization for the general quartic (single literal coefficients). He then re-inserts the compound coefficient terms in the resulting auxiliary cubic, to obtain for this specific problem a ruler-and-compass constructible factorization, which is not available for the general quartic. Only thus can his method match the ancient result he cites, that the problem is ruler-and-compass constructible.

The general quartic here functions as a partial content ("quartic") of the equation of the original problem. It allows one to display more *clearly* the reasons for the ultimate result, even though the outcome depends on special features of the equation studied. First, factorization theory is able to proceed more uniformly by operating on a separate representation of the partial content. The uniformity "brings out" the connection between that partial content and the auxiliary cubic. Second, although the factorization we get after re-inserting the complex coefficient terms into the auxilliary cubic is special to the original equation, obtaining it by the two-tier procedure rather than by some direct argument gives us considerable insight into the reasons for the factorization: the two-tier procedure "brings out" that the ruler-and-compass solution results from the problem being quartic, but of a special type. A direct algebraic factorization of the original equation gives no corresponding insight.

The availability of partial contents in a representational type controls its potential for "bringing out", its potential for the kinds of insights, that arise precisely by exploiting partial contents. Subject matters may made more intelligible by manufacturing (and then exploiting) suitable additional partial contents. This may be accomplished by a suitable change of representation in a modified practice. The representational types must be *content-linked* via a correlating practice: one that provides a cross-representational type response that can be taken to re-express content. For example, Cartesian practice licenses a response to a geometrical problem by an equation, and gives a way of bringing the roots of the equation to bear on the original geometrical problem. Poncelet's practice gives certain ways of augmenting a geometrical diagram in the theory of conics, and modified standards for making attributions to the diagram.

We thus extend the relation of partial content across content-linked representational types. Representations in the second content-linked representational type may function as partial contents of those representations of the first to which they are linked (or vice versa). A Cartesian equational representation of a geometric problem functions as a partial content of the diagram and the discursive text combined. So does a Poncelet-style reading of the original Euclidean formulation. Because different representational types apparently often have radically different repertoires of partial contents, their content-linked use opens up radically different possibilities for understanding.

What we have called the granularity or grain of a representational type is its repertoire of partial contents. It is evidently a rather basic feature of an intellectual practice. Treating human cognitive states as uniformly individuated according to ordinary language representation seems to miss genuine alternative possibilities for human understanding and expression. Treating proofs as uniformly individuated by reconstruction in some formal theory seems to miss the intellectual point of much of the development of geometry. That intellectual content perhaps does not come in representation-independent atoms or propositions is just one of the possible consequences.

References

Apollonius. 1987. *On Cutting off a Ratio. E. Macierowski, transl.* Annapolis: Golden Hind Press.

Barwise, J., and J. Perry. 1983. *Situations and Attitudes.* Cambridge, Massachusetts: MIT Press.

Carnot, Lazare. 1803. *Géométrie de Position.* Paris: Duprat.

Descartes, René. 1954. *La Géométrie. Transl. M. Latham and D. Smith.* Mineola, NY: Dover Publ. Originally, Open Court Publ., 1925. (Page references are to the included facsimile of the original 1637 edition.).

Euclid. 1956. *The thirteen books of Euclid's Elements. Translated with introduction and commentary by Thomas Heath. 3 vols.* Minneola, NY: Dover Publ. Reprint of Second edition, originally Cambridge University press, 1925.

Hilbert, David. 1959. *The Foundations of Geometry. Transl. E. J. Townsend, Reprint ed.* La Salle, Ill.: Open Court Pub. Co. Translated from the 1899 original, with the additions made by the author in the French translation, Paris, 1901, incorporated.

Kripke, Saul. 1979. A Puzzle about Belief. In *Meaning and Use*, ed. A. Margalit. Dordrecht: Reidel.

Leibniz, G. W. 1969. *Philosophical Papers and Letters. Transl. Loemker.* Dordrecht: Reidel.

Manders, Kenneth. 1989. Domain extension and the philosophy of mathematics. *Journal of Philosophy* 86:553–62.

Mueller, Ian. 1981. *Philosophy of mathematics and deductive structure in Euclid's "Elements"*. MIT Press.

Poncelet, Jean-Victor. 1822. *Traité des propriétés projectives des figures...* Paris: Bachelier. Expanded 2^{nd} edition, 2 vols., Gauthier-Villars 1865.

Poncelet, Jean-Victor. 1862–64. *Applications d'analyse et de Géométrie qui ont servi de principal fondement au traite des proprietes projectives des figures. 2 vols.* Paris: Gauthier-Villars.

Proclus. 1970. *A Commentary on the first Book of Euclid's "Elements." G. Morrow, ed. and transl.* Princeton UP. (Page references are to Proclus' text, indicated in the margin).

Tarski, Alfred. 1959. What is elementary geometry? In *The Axiomatic Method*, ed. L. Henkin, P. Suppes, and A. Tarski. 16–29. Amsterdam: North-Holland.

Tropfke, Johannes. 1922–1924. *Geschichte der Elementar-Mathematik, 7 vols.* Berlin/Leipzig: Walter de Gruyter. 2nd edition.

27

Constraints on Coalgebras

KUNIAKI MUKAI

Introduction

Barwise (1989) gives a note on the hereditary subset relation $\sqsubseteq$ on sets viewed as information subsumption and formalizes unification in the universe of hypersets assuming the anti-foundation axiom (Aczel 1988). In his paper, he proves a theorem called the unification lemma which gives a certain characterizing condition for the system of subsumption constraints to be solvable. By the lemma, we can prove the existence of a solution to the constraints like this:

$$\begin{aligned} x &\sqsubseteq \{x, y\} \\ y &\sqsubseteq \{x\}. \end{aligned}$$

Aczel and Mendler (1989) proves the final coalgebra theorem in ZFC^- (without the foundation axiom) that for every set-based functor, there is a final coalgebra for the functor. Here is a simple example of applications of the theorem. Take the following system of equations on sets:

$$\begin{aligned} x &= \{y\} \\ y &= \{x\}. \end{aligned}$$

This system of equations is a coalgebra for the set-based functor *pow*, the power class functor. So by the theorem there is a unique homomorphism from the coalgebra into the final one, which is a solution to the system.

The final coalgebra theorem is conceptually simple but very powerful. For example, the final coalgebra of *pow* is a model of ZFC^- plus the anti-foundation axiom (Aczel 1988). Moreover, there are many useful examples

The author would like to thank Larry Moss for his useful comments on this paper. Diagrams in this paper are drawn with Paul Tailor's commutative diagram style macro `diagrams.tex`.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

which can be viewed as coalgebras for set-based functors: automata, rational trees, directed graphs, feature structures, terms, process, and so on.

Now the objective of the present paper is to generalize Barwise's unification lemma with the final coalgebra theorem to include those data types above. We aim at the extension so that the unification lemma becomes a special case with the data type of hypersets which are generated by the parameter functor *pow*.

Following Aczel and Mendler 1989, we work in the category of classes of sets and functions in ZFC^-. The notion of a coalgebra for an endofunctor is basic. For the purpose of this paper, endofunctors and coalgebras for an endofunctor may be explained informally as follows. First, given a collection of indeterminates, the endofunctor returns a collection of possible forms made up of the indeterminates as components. A coalgebra for the endofunctor is a special kind of a system of equations between indeterminates and possible forms. Each equation in the system gives a possible form to a unique indeterminate of the equations.

Here are basic assumptions on which we put a foundation of constraint problems.

1. There is a final coalgebra for the given endofunctor on the category.
2. Each coalgebra is given a collection of relations on it, which we call subsumptions on the coalgebra.
3. There is the maximum subsumption on the final coalgebra in the given collection.
4. Homomorphisms between coalgebras preserve subsumptions.
5. The restriction of relations to subcoalgebras preserves subsumptions.

For condition (2), we introduce a notion of a relational operator which gives a collection of subsumptions for each coalgebra by some compatibility condition. The relational operator is a kind of function which transforms a relation on indeterminates to a relation on possible forms made up of the indeterminates. We prove in this paper that for any relational operator and coalgebra, there is the maximum (generalized) subsumption on the coalgebra. Our notion of relational operator is so general that it includes all useful examples mentioned above. Furthermore we show that if the relational operator preserves a preorder then the maximum subsumption is also a preorder.

Based on these results, we formulate constraint problems as follows. A constraint problem consists of basic constraints on indeterminates and possible forms of some unknown coalgebra. Basic constraints are divided into two types: equational constraints and subsumption constraints.

A solution to the constraint problem consists of a construction of the unknown coalgebra and a homomorphism from it into the final coalgebra which respect to all the given constraints. More precisely, the solution homomorphism sends both sides of a basic equational constraint to the same element, and of a basic subsumption constraint to a pair of elements which stand in the maximum subsumption relation on the final coalgebra.

Dörre (1990) solved an open problem on feature logic with weak subsumption constraints applying well-known method to transform non-deterministic automata into the deterministic one. We will show an alternative proof as a demonstration our version of unification. Our proof use only obvious properties of the underlying functor for feature structures.

The present work generalizes Mukai (1991), which proposes a constraint logic programming language over hypersets based on Aczel's anti-foundation axiom (AFA). The present paper does not assume the anti-foundation axiom, but work in ZFC^- based on the final coalgebra theorem instead.

27.1 Preliminaries

V denotes the class of sets in ZFC^-. $\mathcal{C}$ is the superlarge category with objects all subclasses of V and with arrows all functions between them. The arrow composition law of $\mathcal{C}$ is the standard functional composition. With functions $f: X \to Y$ and $g: Y \to Z$, we write gf for the functional composition of f and g, namely, $fgx = f(g(x))$ for $x \in X$.

A (covariant) endofunctor Φ on $\mathcal{C}$ is *set-based* if for each class X and $a \in \Phi X$ there is some *set* $X_0 \in V$ and $d \in \Phi X_0$ such that $X_0 \subseteq X$ and $a = \Phi\iota_{X_0,X} d$, where $\iota_{X_0,X}$ is the inclusion map from X_0 into X.

An ordered pair (X, α) of a class X and a function $\alpha: X \to \Phi X$ is called a *coalgebra* for Φ. It is also called a Φ-coalgebra. Given two coalgebras (X, α) and (Y, β) for Φ, f is a *homomorphism* from (X, α) into (Y, β) if the following square commutes for a function $f: X \to Y$:

$$\begin{tikzcd}
\Phi X \arrow[r, "\Phi f"] & \Phi Y \\
X \arrow[u, "\alpha"] \arrow[r, "f"'] & Y \arrow[u, "\beta"']
\end{tikzcd}$$

A coalgebra (A_0, α_0) is a *subcoalgebra* of (A, α) if $A_0 \subseteq A$ and the inclusion map $\iota_{A_0,A}$ is a homomorphism $\iota_{A_0,A}\colon (A_0,\alpha_0) \to (A,\alpha)$.

$$\begin{tikzcd} \Phi X_0 \arrow[r, "\Phi\iota_{A_0,A}"] & \Phi X \\ X_0 \arrow[u, "\alpha_0"] \arrow[r, "\iota_{A_0,A}"'] & X \arrow[u, "\alpha"'] \end{tikzcd}$$

The subcoalgebra is *small* if A_0 is a set. The following lemma are due to Aczel and Mendler 1989.

Lemma 1 (The Small Subcoalgebra Lemma) *If (A,α) is a coalgebra and X is a subset of A then $X \subseteq A_0$ for some small subcoalgebra (A_0,α_0) of (A,α).*

Also we use this lemma to show the existence of the maximum relation on a coalgebra with respect to the given relational operator.

A *final coalgebra for* Φ is a coalgebra (Y,β) for Φ if for any coalgebra (X,α) for Φ, there is a unique homomorphism from (X,α) into (Y,β). Aczel and Mendler (1989) proves the final coalgebra theorem:

Theorem 2 (The Final Coalgebra Theorem) *Every set-based functor on the category C of classes has a final coalgebra for the functor.*

We write $R\colon X \rightharpoonup Y$ for a relation R between a class X and Y: $R \subseteq X \times Y$. A *relation R on X* is a relation $R \subseteq X^2$. If $f\colon X \to Y$ is a function and S is a relation on Y, then $f^*S = \{(a,b) \in X^2 \mid (fa,fb) \in S\}$, which is a relation on X. Also if R is a relation on X, then $f_*R = \{(fx,fy) \mid (x,y) \in R\}$, which is a relation on Y.

$$\begin{tikzcd} Y \arrow[r, rightharpoonup, "S"] & Y & & Y \arrow[r, rightharpoonup, "f_*R"] & Y \\ X \arrow[u, "f"] \arrow[r, rightharpoonup, "f^*S"'] & X \arrow[u, "f"'] & & X \arrow[u, "f"] \arrow[r, rightharpoonup, "R"'] & X \arrow[u, "f"'] \end{tikzcd}$$

27.2 Subsumption Relation on a Coalgebra

The following definition is most basic in this paper.

Definition 1 *A* relational operator *for an endofunctor Φ is a family Δ of operators Δ_X indexed by the subclasses X of V which assigns to each binary relation R on X a binary relation $\Delta_X R$ on ΦX such that the following hold for any mapf: $X \mapsto Y$.*

1. Δ *is* monotone*:* $R \subseteq R' \Longrightarrow \Delta_X R \subseteq \Delta_X R'$ *for any relations* R, R' *on* X*.*
2. Δ preserves inverses*:* $R = f^* S \Longrightarrow \Delta_X R = (\Phi f)^* \Delta_Y S$ *for any relations* R *on* X *and* S *on* Y*.*

$$\begin{tikzcd} Y \arrow[r, "S"] & Y \\ X \arrow[u, "f"] \arrow[r, "R"'] & X \arrow[u, "f"'] \end{tikzcd} \qquad \begin{tikzcd} \Phi Y \arrow[r, "\Delta_Y S"] & \Phi Y \\ \Phi X \arrow[u, "\Phi f"] \arrow[r, "\Delta_X R"'] & \Phi X \arrow[u, "\Phi f"'] \end{tikzcd}$$

We abuse Δ for Δ_X if the context is clear.

Remark We do not assume that relational operators preserve relational compositions.

Example 1 *Given a class* X *and a relation* R *on* X*, define a relation* $\Delta_X^{pow} R$ *on pow* X *so that*

$$(u, v) \in \Delta_X^{pow} R \iff \forall x \in u \exists y \in v \; (x, y) \in R.$$

Then, clearly, Δ_X^{pow} *is monotone and preserves inverses. Hence the mappings* Δ_X^{pow} *indexed by classes* X *form a relational operator for pow. We write* Δ^{pow} *for this relational operator.*

Remark Aczel and Mendler (1989) defines the notion of a *congruence*, which we do not repeat here, but will give an equivalent one later. In fact, it is not difficult to see the equivalence: Given a Φ-coalgebra (X, α), a relation R on X is a congruence in their sense if and only if $R = (\stackrel{f}{\sim} \restriction X)$ for some Φ-coalgebra (Y, β) and a homomorphism $f: (X, \alpha) \to (Y, \beta)$. The symbol $\stackrel{f}{\sim}$ will be defined in definition 3.

Definition 2 *Given a relational operator* Δ *for* Φ *and a coalgebra* (X, α) *for* Φ*, a relation* R *on* X *is* Δ-compatible *on* (X, α) *if* $R \subseteq \alpha^* \Delta R$*, namely,* $(x, y) \in R$ *implies* $(\alpha x, \alpha y) \in \Delta R$ *for any* $x, y \in X$*.*

$$\begin{tikzcd} \Phi X \arrow[r, "\Delta R"] & \Phi X \\ X \arrow[u, "\alpha"] \arrow[r, "R"'] & X \arrow[u, "\alpha"'] \end{tikzcd}$$

Fixing a relational operator Δ in the rest of the paper, we often omit the prefix Δ- when it is clear from the context.

From the definition of a relational operator, we can prove the following lemma that the restriction operation on relations preserves their compatibility.

Lemma 3 (The Restriction Lemma) *Let (A_0, α_0) be a subcoalgebra of (A, α) and R an compatible relation on (A, α). Then the restriction $R \restriction A_0$ is a compatible relation on (A_0, α_0).*

Proof. Let $\iota: A_0 \to A$ be the inclusion map and define $R_0 = R|A_0$ Then $R_0 = \iota^* R$ by definition of $(\ )^*$. As Δ is a relational operator $\Delta R_0 = (\Phi\iota)^* \Delta R$. Assume that $(x, y) \in R_0$. As R is compatible on (A, α), we have $(\alpha x, \alpha y) \in \Delta R$. As (A_0, α_0) is a subcoalgebra of (A, α) we have $((\Phi\iota)\alpha_0 x, (\Phi\iota)\alpha_0 y) \in \Delta R$. Hence, from $\Delta R_0 = (\Phi\iota)^* \Delta R$, we have $(\alpha_0 x, \alpha_0 y) \in \Delta R_0$. This concludes that R_0 is compatible relation on (A_0, α_0). The following cube helps the proof be clear.

$$
\begin{tikzcd}
\Phi A_0 \arrow[rr, "\Delta R_0"] \arrow[dr, "\Phi\iota"] & & \Phi A_0 \arrow[dr, "\Phi\iota"] & \\
& \Phi A \arrow[rr, "\Delta R"] & & \Phi A \\
A_0 \arrow[uu, "\alpha_0"] \arrow[rr, "R_0"] \arrow[dr, "\iota"] & & A_0 \arrow[uu, "\alpha_0"] \arrow[dr, "\iota"] & \\
& A \arrow[uu, "\alpha"] \arrow[rr, "R"] & & A \arrow[uu, "\alpha"]
\end{tikzcd}
$$

□

The following lemma asserts that the image of a compatible relation under a homomorphism is also compatible.

Lemma 4 (The Image Relation Lemma) *Let Δ be a relational operator for Φ, $f\colon (A,\alpha) \to (B,\beta)$ a homomorphism, and R a compatible relation on (A,α). Then the image relation f_*R of R under f is compatible on (B,β).*

Proof. Assume that $(x,y) \in f_*R$. As $R \subseteq f^*f_*R$ by definition of f_*R, we have $\Delta R \subseteq \Delta f^*f_*R = (\Phi f)^*\Delta f_*R$. Moreover, as $f\colon (A,\alpha) \to (B,\beta)$ is a homomorphism, we have $\beta x = (\Phi f)\alpha a$ and $\beta y = (\Phi f)\alpha b$ for some $a, b \in A$. As R is compatible on (A,α) and $\Delta R \subseteq (\Phi f)^*\Delta f_*R$, we have $(\beta f a, \beta f b) \in \Delta f_*R$, which concludes the lemma. The following cube helps the proof be clear.

$$
\begin{tikzcd}
\Phi A \arrow[rr, "\Delta R"] \arrow[dr, "\Phi f"'] & & \Phi A \arrow[dr, "\Phi f"] & \\
& \Phi B \arrow[rr, "\Delta f_* R"] & & \Phi B \\
A \arrow[uu, "\alpha"] \arrow[rr, "R"] \arrow[dr, "f"'] & & A \arrow[uu, "\alpha"'] \arrow[dr, "f"] & \\
& B \arrow[uu, "\beta"] \arrow[rr, "f_* R"] & & B \arrow[uu, "\beta"']
\end{tikzcd}
$$

□

Lemma 5 (Maximum Compatible Relation Lemma) *If Φ is a set-based endofunctor on the category $\mathcal{C}$ of classes and Δ a relational operator for Φ, then for any Φ-coalgebra (X, α), there is the maximum Δ-compatible relation on (X, α).*

Proof. This proof follows basically along the proof proved by Aczel and Mendler (1989) which proves the existence of the maximum congruence relation for any coalgebra for a set-based functor. So we give only the outline of our proof here.

1. Take the union R of all small and Δ-compatible relations on the given coalgebra.
2. Show that R is also Δ-compatible on the coalgebra.
3. Show that any Δ-compatible relation on the coalgebra is a subclass of R.

The above second step is obvious as the relational operator is monotone by definition. In the third step above, with S any Δ-compatible relation on (X, α) and $(x, y) \in S$, we apply the small subcoalgebra lemma 1 to obtain a small subcoalgebra (X_0, α_0) of (X, α) so that $\{x, y\} \subseteq X_0$. By applying the restriction lemma 3 to the subcoalgebra we have a Δ-compatible relation $S|X_0$ on (X_0, α_0). Then applying the image relation lemma 4 to $S|X_0$, we have that $S|X_0$ is small and Δ-compatible on (X, α). Hence we have $S|X_0 \subseteq R$ and $(\alpha x, \alpha y) \in \Delta(S|X_0)$. Since Δ is monotone, we have $(\alpha x, \alpha y) \in \Delta R$, which concludes the lemma. □

Corollary 6 (Maximum Compatible Preorder) *Let Φ be a set-based endofunctor on the category $\mathcal{C}$ of classes and Δ a relational operator for Φ. If Δ preserves preorders then for any Φ-coalgebra (X, α), there is the maximum Δ-compatible preorder on (X, α).*

Proof. By the maximum compatible relation lemma 5, there is the maximum compatible relation R on (X, α) for Φ. We prove that R is also a preorder. Let R' be the reflexive and transitive closure of R: R' is the least preorder such that $R \subseteq R'$. As Δ preserves preoreders, $\Delta R'$ is also a preorder. Hence $(\Delta R')^n \subseteq \Delta R'$ for any integer $n \geq 0$, where $(\Delta R')^0 = \{(x, x) \mid x \in X\}$. As Δ is monotone and $R \subseteq R'$, we have $(\Delta R)^n \subseteq (\Delta R')^n$. Now suppose $(x, y) \in R'$. By definition of R', $(x, y) \in R^n$ for some natural number n. As R is Δ-compatible, $(\alpha x, \alpha y) \in (\Delta R)^n$. Hence $(\alpha x, \alpha y) \in \Delta R'$. Therefore R' is a Δ-compatible relation on (X, α). Hence $R' \subseteq R$. Therefore $R = R'$. □

Example 2 *Applying this theorem to the endofunctor pow and the relational operator Δ^{pow} defined above, we have the maximum Δ^{pow}-compatible relation on the final pow-coalgebra It should be clear that this relation is a model of the hireditary subset relation on the universe of non-well-founded sets.*

Let (A, α) be a final coalgebra for Φ. Note that α is necessarily an injection. Let X be a class, and (R, S) an ordered pair of relations on $X + \Phi X$. Namely, $R, S \subseteq (X + \Phi X)^2$. We call such (R, S) a (Φ, Δ)-*constraint problem over* X.

Now we define the notion of a solution. For generality, we do not assume $X \cap \Phi X \neq \emptyset$ in the following definition.

To define the notion of a solution, we need two auxiliary relations called congruence and simulation on $X + \Phi X$ (direct sum) as follows. Given a homomorphism $f: (X, \alpha) \to (Y, \beta)$, let E be the equivalence relation on the direct sum $W = X + Y + \Phi X + \Phi Y$ generated by $\alpha \cup \beta \cup f \cup \Phi f$, where α, β, f, and Φf are viewed as relations on the sum W in the obvious way. Put $E_0 = E|(X + \Phi X)$ and let $k: X + \Phi X \to X \cup \Phi X$ be the standard surjection and define $\stackrel{f}{\sim} = k_* E_0$. Clearly, $\stackrel{f}{\sim}$ is an equivalence relation on $X + \Phi X$. Here is a trivial example of E and E_0.

Example 3 *Let $X = Y = \{a, b\}$, $\Phi = I$ and $f = id_X$, where $a \neq b$, and I is the identity functor on $\mathcal{C}$. Then*

$$\begin{aligned} W &= X + \Phi X + Y + \Phi Y = X + X + X + X \\ &= \{(1, a), (1, b), (2, a), (2, b)(3, a), (3, b), (4, a), (4, b)\}, \\ W/E &= \{\{(1, a), (2, a), (3, a), (4, a)\}, \{(1, b), (2, b), (3, b), (4, b)\}\}, \end{aligned}$$

and

$$(X + \Phi X)/E_0 = \{\{(1, a), (2, a)\}, \{(1, b), (2, b)\}\}.$$

Definition 3 (Congruence) *With* $\overset{f}{\sim} = k_* E_0$ *as above, a relation* R *on* $X + \Phi X$ *is a* congruence *on* (X, α) *if* $R = \overset{f}{\sim}$ *for some homomorphism* f.

This definition of a congruence is essentially the same as in Aczel and Mendler (1989).

Next, for a relation S on Y, define

$$f^+ S = E_0(f^* S \cup (\Phi f)^* \Delta S) E_0,$$

where the right-hand side is a relational composition of relations on $X + \Phi X$. Note that $f^+ S$ is a relation on $X \cup \Phi X$.

Definition 4 (Simulation) *A relation* S *on* $X \cup \Phi X$ *is a* simulation *if* $S = id_X^+ S'$ *for some* Δ*-compatible relation* S' *on some* Φ*-coalgebra* (X, α), *where* id_X *is the identity homomorphism* $id_X \colon (X, \alpha) \to (X, \alpha)$.

Definition 5 (Solution) *A* Φ*-coalgebra* (X, α) *is a* solution *to a given* (Φ, Δ)*-constraint problem* (R, S) *over* X *if there is a homomorphism* f *from* (X, α) *into a final* Φ*-coalgebra such that* $R \subseteq \overset{f}{\sim}$ *and* $S \subseteq f^+(\sqsubseteq)$, *where* $\sqsubseteq$ *is the maximum* Δ*-compatible relation on the final coalgebra.*

Lemma 7 (Epi-Mono Decomposing Lemma) *Any morphism* $f \colon (X, \alpha) \to (Y, \beta)$ *may be written* $f = hg$ *for some coalgebra* (Z, γ) *so that* $g \colon (X, \alpha) \to (Z, \gamma)$ *is surjective and* $h \colon (Z, \gamma) \to (Y, \beta)$ *is injective.*

Proof. Thc proof is easy. □

This lemma is important from computational view point. Assuming some computability conditions on Φ and Δ, if an exhaustive method is available for searching for all surjective homomorphisms from Φ-coalgebras on a given finite set, then (Φ, Δ)-constraint problems are decidable.

Theorem 8 (Main Theorem) *If* Φ *is a set-based functor,* Δ *a relational operator for* Φ, *and* (R, S) *a* (Φ, Δ)*-constraint problem over a class* X, *then the problem has a solution if and only if* $R \subseteq R'$ *and* $S \subseteq S'$ *for some congruence* R' *and simulation* S' *on* $X \cup \Phi X$.

Proof. (only if part): Suppose (X, α) is a solution and f is the unique homomorphism from (X, α) into the final Φ-coalgebra. Then the definition of a solution gives $R \subseteq \overset{f}{\sim}$ and $S \subseteq f^+(\sqsubseteq)$. Clearly $\overset{f}{\sim}$ is a congruence and $f^+(\sqsubseteq)$ is a simulation.
(if part): Suppose $R \subseteq R'$ and $S \subseteq S'$ for some congruence R' and simulation S' on $X \cup \Phi X$. Then by definition of a congruence and a simulation, $R' = \overset{f}{\sim}$ and $S' = id_X^+ S''$ for some $\alpha \colon X \to \Phi X$, Φ-coalgebra (Y, β), and $f \colon (X, \alpha) \to (Y, \beta)$. Let (Z, γ) be the final Φ-coalgebra, and $g \colon (Y, \beta) \to (Z, \gamma)$ the unique homomorphism. Clearly, $\overset{f}{\sim} \subseteq \overset{gf}{\sim}$. Hence $R \subseteq \overset{gf}{\sim}$. Also $id_X^+ S'' = f^*(f_* S'')$ is clear. By the image lemma 4, $f_* S''$

is compatible on (Y, β). So $f_* S'' \subseteq g^+(\sqsubseteq)$, where $\sqsubseteq$ is the maximum Δ-compatible relation on the final coalgebra. Hence $S \subseteq (gf)^+(\sqsubseteq)$, which concludes the if-part.

$$\begin{tikzcd} \Phi X \arrow[r, "\Phi f"] & \Phi Y \arrow[r, "\Phi g"] & \Phi Z \\ X \arrow[u, "\alpha"] \arrow[r, "f"'] & Y \arrow[u, "\beta"'] \arrow[r, "g"'] & Z \arrow[u, "\gamma"'] \end{tikzcd}$$

$\square$

Example 4 *Clearly, the identity class functor I on $\mathcal{C}$ is set-based It is not difficult to see that for any I-coalgebra (X, γ), X is a final I-coalgebra if and only if X is a singleton. Hence V is not a final coalgebra for the functor, though V is the maximum fixpoint of I. Then for any class X, Φ-constraint problem over X, a coalgebra (X, β) is a solution to the problem for any $\beta: X \to X$. The unique homomorphism from (X, β) into the singleton is the obvious mapping.*

27.3 Examples of Constraint Problems

Here are two examples of our framework of constraints. One is feature structure constraints, and the other is Barwise's unification lemma.

Example 5 (Feature Structure) *Let A be a finite set. Let M_A be an endofunctor on $\mathcal{C}$ such that $M_A X$ is the class of partial functions from A into X and that for $f: X \to Y$, $M_A f$ is defined by the conditions $dom(M_A f r) = dom\, r$ and $M_A f r a = f(ra)$ for $a \in dom(M_A f r)(= dom\, r)$. In fact, M_A is an endofunctor. Next, define a relational operator Δ_A for M_A by*

$$(r, s) \in \Delta_A R \iff dom\, r \subseteq dom\, s \wedge \forall a \in dom\, r\, (ra, sa) \in R.$$

It is easy to check that Δ_A is a relational operator for M_A which preserves preorders.

It is obvious from the definition of M_A that if X is finite then $M_A X$ is also finite. So by the remark after lemma 7, (M_A, Δ_A)-constraint problem over a finite number of indeterminates is decidable. It should be clear that (M_A, Δ_A)-constraint problems are external feature structure constraint problems which was first solved by Dörre 1990[1].

[1]The intensional version of feature structure constraint problems is undecidable (Dorre and Rounds 1990)

Example 6 (Defining Merge Operation) *Let (X, α) is a Φ-coalgebra. Assume that $(pow\, X, \beta)$ is a coalgebra such that the diagram commutes for some β and γ:*

$$\begin{tikzcd}
\Phi(pow\, X)^2 \arrow[r, "\Phi\cup"] & \Phi(pow\, X) & \Phi X \arrow[l, "\Phi j"'] \\
(pow\, X)^2 \arrow[u, "\gamma"] \arrow[r, "\cup"'] & pow\, X \arrow[u, "\beta"'] & X \arrow[u, "\alpha"'] \arrow[l, "j"]
\end{tikzcd}$$

where $j \colon X \to pow\, X$ is defined by $jx = \{x\}$.

Let p and q be the unique homomorphisms from $pow\, X$ and $(pow\, X)^2$ into the final coalgebra, respectively. The image of the coalgebra $(pow\, X, \beta)$ under q form a subcoalgebra (D, δ) of the final coalgebra. Define a binary operation $\oplus$ on D by $\oplus(px, py) = q\langle x, y\rangle$ for $x, y \in pow\, X$. Clearly from the above diagram, $\oplus$ is well-defined. Also from the diagram and properties of the set union operator $\cup$ on $pow\, X$, it is easy to see that the following equations hold of the subcoalgebra (D, δ).

$$\begin{aligned}
(a \oplus b) \oplus c &= a \oplus (b \oplus c) \\
a \oplus b &= b \oplus a \\
a \oplus a &= a \\
\varepsilon \oplus a &= a
\end{aligned}$$

where $\varepsilon = p\emptyset$. Thus we have obtained an algebra which behaves in a way similar to that of feature structures, which is associative, commutative, and idempotent with a unit. If X is finite, this algebra is finite. For the case in which $\Phi = M_A$, we can give the above β and γ which together correspond to the well known procedure to merge states of an automaton. So we can give effectively a non-exhaustive search algorithm for (M_A, Δ_A)-constraint problems.

We show a version of the unification lemma proved by Barwise (1989). To do so, we first introduce "atoms" and "indeterminates" without expanding the universe V, which is unlike Barwise (1989). Given a class X, we denote the class $\{(0, x) \mid x \in X\}$ by $X^\natural$. An element of $V^\natural$ is called an indeterminate. Let A be a class which has no indeterminate and $\mathcal{D}$ the full subcategory of $\mathcal{C}$ which has all classes of indeterminates as objects. There is an obvious isomorphic functor Ψ from $\mathcal{C}$ onto $\mathcal{D}$. Define a functor pow'_A from $\mathcal{D}$ into $\mathcal{C}$ by $pow'_A X = pow(A \cup X^\natural)$ and $(pow'_A f)u = (u \cap A) \cup \{fx \mid x \in u \setminus A\}$, where $f \colon X \to Y$ and $u \subseteq A \cup X$. As A has no indeterminate, pow'_A is well-defined and is a functor from $\mathcal{D}$

into $\mathcal{C}$. We call the endofunctor $pow_A = (pow'_A)\Psi$, a *power class functor*. Clearly, pow_A is a set-based functor on $\mathcal{C}$.

Remark If in the above we define $pow'_A fu = \{fx \mid x \in u\} \cup ((u \cap A) \setminus X)$ then pow'_A is not a functor. To show this, let $A = \{a, b\}$, $f: \{x\} \to \{a, x\}$, and $g: \{a, x\} \to \{b\}$. Then it is checked that $(pow'_A gf) \neq (pow'_A g)(pow'_A f)$, because $\{a, x\}$ is sent to $\{a, b\}$ by the left-hand side and to $\{b\}$ by the right-hand side. Hence pow'_A is not a functor.

We use the symbol pow_A^+ for the relational operator defined by

$$(u, v) \in pow_A^+ R \iff A \cap u \subseteq A \cap v \wedge \forall (0, x) \in u \cap X^\natural \exists (0, y) \in u \cap X^\natural \ (x, y) \in R,$$

where R is a relation on X.

Example 7 *Now take a (pow_A, pow_A^+)-constraint problem (E, S), where $E = \{(x, \{a, y\}), (y, \{x\})\}$, $S = \{(\{x, y\}, \{x\})\}$, $A = \{a, b\}$, $X = \{x, y\}$. The problem may be written as the usual form.*

$$\begin{aligned} x &= \{a, y\} \\ y &= \{x\} \\ \{x, y\} &\sqsubseteq \{x\} \end{aligned}$$

The simple exhaustive search shows that there is no simulation which extend S. So by the main theorem 8, this constraint problem has no solution.

Definition 6 *Let A be a class of non-indeterminates, namely, $A \cap V^\natural = \emptyset$. By a* bisimulation *we mean an equivalence relation $\sim$ on $X \cup pow_A X$ satisfying the following condition. If $u \sim v$, then:*

1. *If $u \in A$ then $u = v$.*
2. *If neither u nor v is an indeterminate, then for every $u' \in u$ there is a $v' \in v$ such that $u' \sim v'$.*
3. *If $x \in X$, then there is some non-indeterminate u such that $x \sim u$.*

By an *anchor* we mean any function from X into the underlying class of a final pow_A-coalgebra.

Definition 7 *By a* simulation pair *we mean a pair of relations $\preceq$ and $\sim$ on $X \cup pow_A X$, satisfying the following conditions:*

1. *$\sim$ is a bisimulation relation.*
2. *$\preceq$ is a simulation relation. That is:*
   - *If $u \preceq v$, and either $u \in A$ or $v \in A$, then $u = v$.*
   - *If $u \preceq v$, and $u, v \in pow_A X$, then for all $u' \in u$ there is a $v' \in v$ such that $u' \preceq v'$.*
3. *$\sim$ is a congruence relation with respect to $\preceq$. That is,*
   - *$u \sim v$ implies $u \preceq v$;*

- *$u \preceq v$, $u \sim u'$, and $v \sim v'$ implies $u' \preceq v'$.*

Theorem 9 (Unification Lemma) *Suppose we are given two indexed families $\{a_i \in X \cup pow_A X \mid i \in I \cup J\}$ and $\{b_i \in X \cup pow_A X \mid i \in I \cup J\}$ with $X \cap pow_A X = \emptyset$. Then the following are equivalent:*

1. *There is an anchor f such that for each $i \in I$, $\hat{f}a_i = \hat{f}b_i$, and for each $j \in J$, $\hat{f}a_j \sqsubseteq \hat{f}b_j$, where, $\hat{f}u = (pow_A f)u$ if $u \in pow_A X$ and $\hat{f}u = fu$ otherwise, and $\sqsubseteq$ is the maximum pow_A-subsumption relation on the final pow_A-coalgebra.*
2. *There is a simulation pair $\preceq$ and $\sim$ satisfying $a_i \sim b_i$ for each $i \in I$ and $a_j \preceq b_j$ for $j \in J$.*

Proof. (1)$\Longrightarrow$(2): Suppose (1). Then clearly we can construct some pow_A-coalgebra (X, α) such that f is a homomorphism. $\overset{f}{\sim}$ and $f^+(\sqsubseteq)$ are clearly a bisimulation and a simulation, respectively, on $X \cup pow_A X$.
(2)$\Longrightarrow$(1): Notice that a bisimulation for the endofunctor pow_A is a congruence and that a simulation in the sense of definition 7 is also a simulation in the sense of definition 4. Hence obviously the main theorem 8 concludes this case.

$\square$

References

Aczel, P. 1988. *Non-well-founded Sets.* CSLI lecture notes Number 14. CSLI Publications, Stanford University.

Aczel, P., and N. Mendler. 1989. A Final Coalgebra Theorem. In *Category Theory and Computer Science*, ed. P.H. Pitts, D.E. Rydeheard, P. Dybjer, A.M. Pitts, and A. Poigné. LNCS, No. 389. Springer-Verlag.

Barwise, J. 1989. *The Situation in Logic.* CSLI Lecture Notes 17. Stanford: CSLI Publications.

Dörre, J. 1990. Feature Logic with Weak Subsumption Constraints. Technical Report IWBS tech. Report 101. IBM Deutschland,Stuttgart, April.

Dörre, J., and W. Rounds. 1990. On Subsumtpion and Semiunification in Feature Algebra. In *Proceedings of Logic in Computer Science.*

Mukai, K. 1991. CLP(AFA): Coinductive Semantics of Horn Clauses with Compact Constraints. In *Situation Theory and Its Applications, volume II*, ed. J. Barwise, G. Plotkin, J.M. Gawron, and S. Tutiya. CSLI Publications, Stanford University.

28

Proof Styles in Multimodal Reasoning

Jon Oberlander, Richard Cox and Keith Stenning

Introduction: Questions of Style

Does multimodal logic teaching create a homogeneous population of human theorem provers? Or do students develop individual styles of proof? If their styles differ, what patterns emerge? Can these patterns be predicted from other information about the students?

Hyperproof is a computer program created by Barwise and Etchemendy for teaching first-order logic. It uses multimodal graphical and sentential methods, and is inspired by a situation-theoretic approach to heterogeneous reasoning (Barwise and Etchemendy 1994). A distinctive feature of Hyperproof is its set of 'graphical' rules, which permit users to transfer information to and fro, between graphical and linguistic modes. We have been carrying out a series of experiments on Hyperproof, to help evaluate its effects on students learning logic.

In earlier work (Stenning and Oberlander 1991, 1995), we have emphasised the idea that graphical systems possess a useful property—over-specificity—whereby certain classes of information must be specified. The property is useful because inference with such specific representations can be very simple. We have also urged that actual graphical systems—such as Hyperproof—do allow abstractions to be expressed, and it is this that endows them with a usable level of expressive power. We are therefore interested in determining empirically how students respond to Hyperproof's abstraction mechanisms.

The plan of this paper is as follows: we introduce Hyperproof, and then study two cases of proofs constructed in Hyperproof. The two stu-

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

dents addressed the same problem, but we observe a number of significant differences in the way they solved it. We then discuss the experimental regime under which these proofs were gathered; we focus on our finding that there is a robust distinction between subjects who are more or less successful on an independent task, whose solution can involve the use of external representations (such as tables). We then return to properties we noted in the case studies, and show how their patterns of rule use and proof structure reflect systematic differences between the two classes of subjects. We conclude by suggesting how these patterns might be explained by the 'specificity hypothesis' we have developed in earlier work.

28.1 The Hyperproof Interface

As can be seen in Figure 1, the interface contains two main windows: one presents a diagrammatic view of a chess-board world containing geometric objects of various shapes and sizes; the other presents a list of sentences in predicate calculus; control palettes are also available. The main windows are used in the construction and editing of proofs. Several types of goals can be proved, involving the shape, size, location, identity or sentential descriptions of objects; in each case, the goal can involve determining some property of an object, or showing that a property *cannot* be determined from the given information. A number of rules are available for proof construction; some of these are traditional syntactic rules (such as $\wedge$-elimination); others are 'graphical', in the sense that they involve consulting or altering the situation depicted in the diagrammatic window. In addition, a number of rules check properties of a developing proof. Hyperproof should be viewed as a proof-checking environment designed to support human theorem proving using heterogeneous information.

In Question 4, shown here, a student is confronted by a graphical situation in which little is known—only that there are three objects of unknown shape or size, side by side—and a set of three linguistic premises—comprising two conditionals relating the shape and size of two objects, and a formula telling us that there are no objects of a certain size. When little is fixed in the graphical situation we term a question *indeterminate* in type, and contrast it with those *determinate* questions in which all the relevant information is specified; we return to these notions in Section 28.3. Here, the student must achieve four proof goals: the first three are shape goals: the students must determine the shapes of the three objects in the world; the last goal is a syntactic goal: the student must determine whether or not a certain formula follows from the graphical and linguistic premises.

28.2 Case Studies

Now consider the two proofs side-by-side in Figure 2. These are responses to

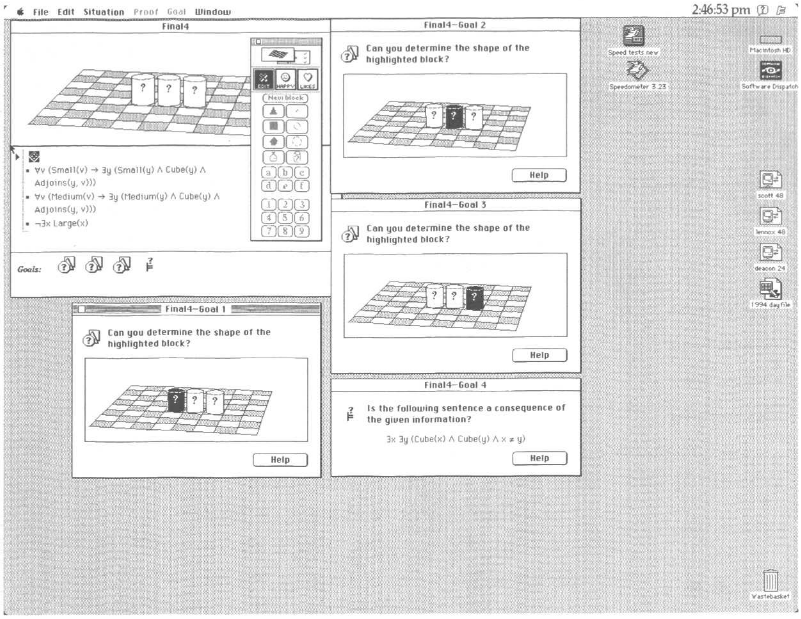


FIGURE 1 The Hyperproof Interface.
The main window (top left) is divided into an upper graphical window, and a lower calculus window. The tool palette is floating on top of the main window, and the other windows reveal a set of goals which have been posed. To achieve them, a proof must be developed, by applying a set of multimodal inference rules to the graphical and calculus premises given.

Question 4, developed by students using Hyperproof under exam conditions. The students (C2 and C14) were both successful with this exam question: they proved all four goals. But they did so with proofs which look somewhat different. There are at least three considerable differences, involving: (i) structural aspects of the proofs; (ii) patterns of rule use; and (iii) treatment of 'graphical variables'. Let us discuss each of these in turn.

28.2.1 Structural Aspects of Proofs

The most obvious difference between C2's proof and C14's is that is shorter: C2 uses just 21 steps to achieve the four goals, while C14 uses 27.[1] A second obvious difference is that C2's proof is shallower, in that there is just one level of embedding in C2's proof (indicated by the heavy bar enclosing user steps 1–14 of C2's proof). C14, by contrast, uses two such levels, and within

[1] These figures include premises.

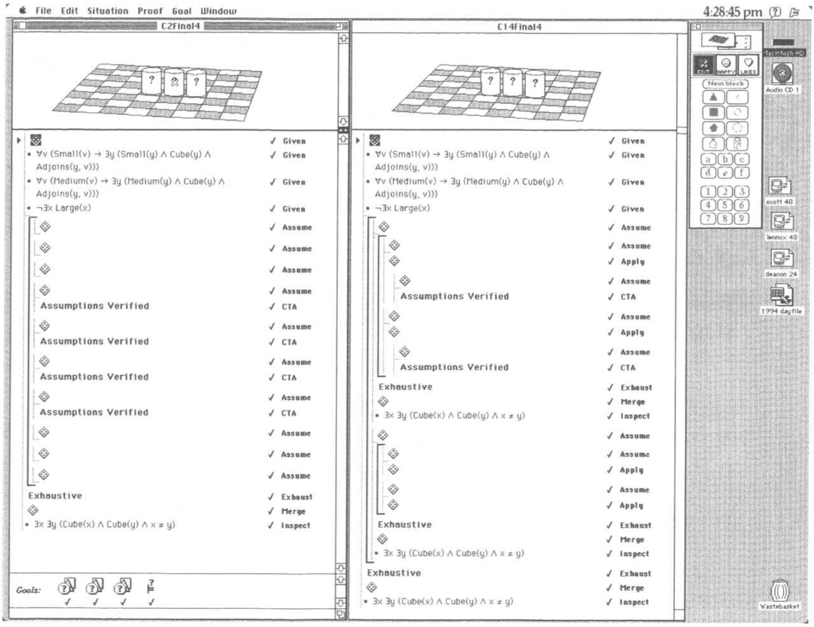


FIGURE 2 Two different subjects' proofs given in answer to an exam question.
The subject on the left (C2) is a 'DetLo' (more verbal in style); the subject on the right (C14) is a 'DetHi' (more diagrammatic in style).

these, uses further embedding (cf. the relation between user steps 3 and 4 of C14's proof). However one measures 'proof depth', it is clear just from inspection that C14's proof is more deeply nested than C2's. There is a third global difference between C2 and C14's proofs, although this one is not apparent from direct inspection of Figure 2. C2's proof was constructed more rapidly than C14's: C2 takes 14 minutes, while C14 takes 19.

So, to summarise, C2's proof in Question 4 is shorter, shallower, and faster.

28.2.2 Patterns of Rule Use

Looking at the proofs in more detail, we find that there are differences in the way that rules have been deployed. Hyperproof's relevant graphical rules are summarised in Table 1. Consider first the pattern that characterises C2's proof: the subproof from steps 1–14 begins and ends with three invocations of Assume; the central part of it possesses a repetitive pattern of Assume, CTA, Assume, CTA, Assume, CTA.

Both C2 and C14 end their proofs with the same final pattern: they

TABLE 1
A set of relevant Hyperproof rules.

RULE	DESCRIPTION
Apply	Extracts information from a set of sentential premises, and expresses it graphically
Assume	Introduces a new assumption into a proof, either graphically or sententially
Inspect	Extracts common information from a set of cases, and expresses it sententially
Merge	Extracts common information from a set of cases, and expresses it graphically
Observe	Extracts information from the situation, and expresses it sententially
Close	Declares that a sentence is inconsistent with either another sentence, or the current graphical situation
CTA	(Check truth of assumptions) Declares that all sentential and graphical assumptions are true in the current situation
Exhaust	Declares that a part of a proof exhausts all the relevant cases

show that they have a exhausted a set of cases, with Exhaust, then show what is common to all of them in a graphical situation, with Merge, and then draw out the linguistic conclusion from that situation, via Inspect.

But the body of C14's proof differs from C2's. It is true that C14 has two instances of an Assume–CTA pattern; but these are essentially parallel structures, each following an instance of an Assume–Apply pattern. Further on in C14's proof, we also find a further pair of Assume and Apply patterns. And while C2 uses the Exhaust–Merge–Inspect pattern just once, at the end of their proof, C14 uses it three times in total, ascending twice from subsubproofs, and once from the larger subproof.

It is clear that these differing patterns fit together with the relative depth of the two proofs: C2 uses repetitive patterns in a shallow proof; C14 avoids the repetition by creating a deeper proof, which in turn requires a particular pattern of rules to recur, as information is drawn together from exhaustive sets of subcases.

In sum, C2 uses Assume–CTA more frequently and repetitively than C14. C14 uses Assume–Apply, and Exhaust–Merge–Inspect more frequently.

28.2.3 Treatment of Graphical Variables

A further difference emerges when one examines the graphical situations which correspond to the diamond-shaped situation icons in the body of the proof. Obviously, C2 uses fewer graphical situations than C14, with 11 as opposed to 15. However, the graphical situations themselves are interestingly different.

Hyperproof's graphical window contains two sorts of symbols, which we may think of as *concrete* and *abstract*. Consider the three symbols that appear in the fifth situation of C14's proof (Figure 3). The righthand symbol is small and cubic, and obviously enough, it depicts a small cube. The central symbol is a small paper bag; however, it doesn't depict a small bag, but rather an object of known (small) size and location, but of unknown shape. It is an abstraction device, in that a picture containing it abstracts over three possible situations, corresponding to the three possible shapes the object could be (cube, tetrahedron or dodecahedron). The lefthand symbol is a cylinder sporting a question mark; it doesn't depict any sort of cylinder, however, but an object of unknown size or shape. Like the paper bag, a cylinder allows us to abstract over several situations. A question-marked cylinder in fact abstracts over nine situations in total. Although not shown here, symbols can also be removed from the checkerboard, and placed by its side, in order to abstract over many possible situations, corresponding to the possible locations the depicted object could occupy. We may therefore contrast concrete symbols (like the small cube) with abstract symbols (like

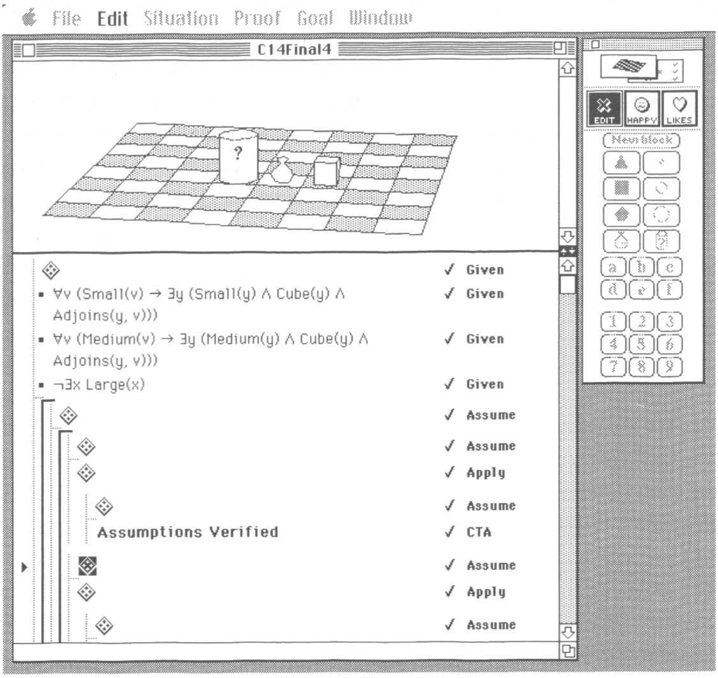


FIGURE 3 Use of graphical abstraction symbols.
The situation being viewed is the fifth in the course of the proof, and contains three symbols of varying degrees of abstraction. The lefthand symbol denotes an object of unknown size or shape; the middle symbol denotes an object of known size but unknown shape; and the righthand symbol denotes an object of known size and shape.

the small bag, or the cylinder). The latter function as graphical variables, more or less.

It should be obvious that some variables are more abstract than others: the question-marked cylinder specifies less information than the small paper bag. Graphical situations containing many highly abstract symbols will, in turn, seem globally more abstract than those which contain symbols which abstract less, or not all. This being so, C2's and C14's proofs can be categorised in terms of the 'abstractness' of the graphical situations they contain. Scanning through the situations, it seems as if C14's are generally more abstract; each contains the same number of graphical symbols as C14's; but in many cases, the symbols are abstract ones, and often highly abstract ones (such as the example in Figure 3).

This first impression of 'extra abstractness' in the case of C14 can be made slightly more precise, by attaching an *index of determinacy* to the graphical situations in the proof. In essence, we can give each graphical symbol in a situation a score: for each attribute (size, shape, location, and label), a symbol scores 1 if that attribute is specified, and 0 otherwise. By totalling the scores, we can give the situation an index: modulo the number of symbols, the higher the score, the more determinate the situation; the lower the score, the more abstract. So, in Figure 3, the lefthand symbol scores 1, the middle symbol 2, and the righthand symbol 3. As a result, the situation as a whole scores 6. Now, if we score each of the situations in each of C2's and C14's proofs, this is the pattern of scores which emerges:

C2 9 9 9 9 9 9 9 9 9 9 4

C14 4 6 7 9 6 7 9 5 4 6 7 7 5 4

Taking the means of these figures, we see that C2's situations have a mean determinacy of 8.55; C14's have a mean of 6.13. There is thus a substantial quantitative gap, which vindicates the initial impression that C14 is entertaining more abstract situations.

28.2.4 Case Studies: Summary

So, in response to Question 4, C2 constructs a shorter, shallower, faster proof, containing Assume–CTA rule cycles, and entertaining situations with relatively little abstractness. C14 constructs a longer, deeper, slower proof, with fewer Assume–CTA cycles, but more Exhaust–Merge–Inspect cycles, and entertaining situations with rather more abstractness.

These cases raise at least three types of question: (i) Are the differences between C2 and C14 accidental, or do they represent differing cognitive styles that we may attribute to groups of individuals? (ii) We have stated that Question 4 is of an indeterminate type; do the kinds of differences we have pointed to also occur on more determinate questions? (iii) Are there underlying reasons linking together the properties in, say, C14's proof?

To answer questions (i) and (ii), we must first describe our experimental setup, and show how we found there to be two distinctive groups among the Hyperproof subjects; we can then indicate in section 28.4 how the proof style differences we noted in our case studies appear to be systematically related to these differences in cognitive style. To answer question (iii), we discuss in section 28.5 how the properties we have drawn attention to may be explained given our cognitive theory.

28.3 Hyperproof Experiments: Method

Two groups of subjects were compared; one group (n = 22) attended a one-quarter duration course taught using the heterogeneous reasoning approach of Hyperproof. A comparison group (n = 13) were also taught for one quarter but in the traditional syntactic manner supplemented with exercises using a graphics-disabled version of Hyperproof (to control for the motivational and other effects of computer-based activities). A fuller description of the method and procedure is provided in Cox et al. 1994.

All subjects were administered two kinds of pre and post-course paper and pencil test of reasoning. The first test was of 'analytical reasoning' and contained two kinds of item derived from the GRE-type of scale of that name (see for example, Duran et al. 1987). We refer to this test as the 'GRE' test. The first GRE subscale consists of verbal reasoning/argument analysis. The other GRE subscale consisted of items often best solved by constructing an external representation of some kind (such as a table or a diagram). We label these subscales as 'indeterminate' and 'determinate', respectively.

The second paper and pencil test we term 'Blocks world'. This test requires reasoning about blocks-world situations like those used in Hyperproof, but is couched in natural language rather than first order logic.

Both groups also sat post-course, computer-based Hyperproof exams. The exam questions differed for the two groups, however, since the syntactic group had not been taught to use Hyperproof's systems of graphical rules. The four questions set the Hyperproof group, though, contained two types of item: determinate and indeterminate. Figures 2 and 3 illustrate Question 4, one of the two indeterminate questions. Student-computer interactions were dynamically logged—this approach might be termed 'computer-based protocol taking'. The logs were time stamped and permitted a full, step-by-step, reconstruction of the time course of the subject's reasoning. The results reported here are based on analyses of those protocols.

Scores on the determinate subscale of the GRE test were used to classify subjects within both Hyperproof and syntactic groups into DetLo and DetHi sub-groups. In other words, the score reflects subjects' facility for solving a type of item that often is best solved using an external representation. DetHi and DetLo subjects in the Hyperproof and syntactic groups responded differently to traditionally versus heterogeneously taught courses; those results are reported in Cox et al. 1994.

Here, however, we are concerned with comparing proof-style differences on the exam questions between DetLo and DetHi subjects' within the Hyperproof group only.

28.4 Systematic Proof Style Differences

We can now return to questions (i) and (ii): are the differences in proof style between C2 and C14 accidental?; and are they as dramatic on determinate questions as on indeterminate ones?

The answer to question (i) should by now be apparent. C2 is a fairly

TABLE 2
Mean parameters of exam proofs:
DetHi/Lo subjects within Hyperproof group.
Score is number of goals achieved (out of 3 in Q1–3; out of 4 in Q4);
Time is in minutes; Steps is steps in overall proof; Depth range is 0–3.

			SCORE	TIME	STEPS	DEPTH	n
DetLo	Det	Q1	3.0	8.42	14.7	0.22	9
		Q3	2.7	18.10	17.2	0.22	9
	Indet	Q2	3.0	11.60	18.6	0.11	9
		Q4	3.5	19.70	27.8	0.56	9
DetHi	Det	Q1	3.0	6.64	17.6	0.23	13
		Q3	3.0	15.14	21.3	0.23	13
	Indet	Q2	2.8	15.23	16.0	0.23	13
		Q4	3.8	20.10	28.6	0.85	13

typical DetLo subject; C14 is a fairly typical DetHi subject. When we take together the performances of all the Hyperproof subjects on the four exam questions, we can uncover some significant results relating to proof parameters, rule usage and use of graphical variables. These results go some way towards showing that the differences between DetLo and DetHi subjects reach down into their styles of proof. Regarding question (ii), it also seems that the differences in proof style are more pronounced in indeterminate exam questions (2 and 4) than in the determinate questions (1 and 3).

28.4.1 Structural Aspects of Proofs

Preliminary analyses were performed on several parameters of these examination proofs. Each proof-log was coded for score (number of proof goals validated), time (time spent on proof), number of proof steps and the proof depth (the depth of nested subproofs the subjects used in their solution). Table 2 shows the mean proof parameters for DetHi and DetLo subjects within the Hyperproof class. There thus seems to be a tendency for DetHi subjects to produce longer, more accurate, and more nested proofs than their DetLo counterparts.

Comparisons between DetLo and DetHi subjects were not statistically reliable, due to wide variation between subjects within the groups. However, taken together, the *pattern* of proof parameters shown in Table 2 suggests superior proof development strategies on the part of Hyperproof DetHi subjects.

In the case studies, we observed that C2's proof in Question 4 was shorter, shallower, and faster. The only uncharacteristic fact about this proof, then, is that a DetLo subject constructed their proof more rapidly than a DetHi; C2, in fact, was one of the fastest DetLo subjects on this question, while C14 was one of the slower DetHi subjects.

28.4.2 Patterns of Rule Use

For the analyses, rule use frequencies for the two indeterminate questions were added and frequencies for the two determinate questions were added.

A two-factor ANOVA for subjects (DetHi, DetLo) and item determinacy (determinate, indeterminate) was conducted separately for each of the rules. The results of these analyses revealed that all subjects used the following rules significantly more frequently[2] in developing proofs for the two indeterminate questions than for the two determinate questions: Assume, Apply and CTA. The Close rule was used significantly more on the *determinate* than on indeterminate questions. A two-way interaction was significant in one of the analyses: the Apply rule was used more on determinate questions by DetLo subjects than by DetHi subjects. Conversely, on indeterminate questions, DetHi subjects used it more frequently than DetLo subjects.

Cluster analyses of the rule use patterns of DetLo and DetHi subjects was also used as an initial exploratory method. These reveal *correlations* between rule uses and suggest the following observations. First, in general, DetLo subjects seem to make CTA a more central part of their rule repertoire than do DetHi subjects, who exploit Exhaust more centrally. Secondly, DetLo subjects seem to have a more stable set of relationships between their rules; the only rule which seems substantially less central for them on indeterminate questions is Close. Thirdly, the more flexible DetHi subjects may use CTA on indeterminate questions more frequently than on determinate questions, but the rule does not correlate closely with their other central rules. By contrast, Apply, and Inspect do seem central, on indeterminate (but not determinate) questions. Finally, like DetLo subjects, DetHi subjects use Close less frequently and centrally on indeterminate questions; however, it remains well correlated with Observe, which one might therefore conclude is also less central a weapon on indeterminate questions.

We observed that C2 used Assume–CTA more frequently and repetitively than C14. C14 used Assume–Apply, and Exhaust–Merge–Inspect more frequently. We can now see that these differences are indeed characteristic of the groups as a whole. We can also see that DetHi subjects, such as C14, have more flexible strategies, and appear to resemble their DetLo colleagues more on determinate questions than on indeterminate questions, like Question 4.

[2]As evidenced by significant main effect for determinacy factor.

28.4.3 Treatment of Graphical Variables

We used determinacy indices to show that C14 is entertaining more abstract situations than C2, on Question 4. Using the indeterminacy index scoring method described in Section 1.3, we can derive scores for all the DetLo and DetHi subjects. So far, we have derived these scores for one of the determinate questions (Question 1) and one of the indeterminate questions (Question 4).

Considering Question 1, all subjects in both the DetLo and DetHi subgroups proved all three proof goals. The index of determinacy scores for DetHi and DetLo subjects proofs did not differ significantly. The mean index of determinacy score for DetLo subjects was $17.6, SD = .31, n = 9$. For the DetHi subjects, the mean was $17.7, SD = .25, n = 13$.

Considering Question 4 (indeterminate), two subjects (one DetLo, one DetHi) did not succeed in proving all of the proof goals. Considering only the subjects who did succeed in proving the proof goals, a one-tailed t-test between DetLo and DetHi subjects index of determinacy scores reveals a significant effect ($t = 1.91, df = 18, p < .05$). The mean index of determinacy score for DetLo was $7.98, SD = .92, n = 8$ and for DetHi it was $7.13, SD = .98, n = 12$. The lower mean index of determinacy score for DetHi indicates more use of abstraction in the steps of the proof.

Thus some support is provided for the prediction, based on specificity theory, that DetHi subjects are more skilled in the deployment of graphical abstraction conventions during reasoning.

28.5 Conclusions

So, C2 and C14 represent two differing cognitive styles; and their differing proofs are characteristic of these styles. They differ in length, depth, patterns of rules used, and quantity of graphical abstraction. But what ties these differences together?

Indeterminate problems, such as Question 4, demand that subjects entertain multiple cases during the course of the proof. There are basically two ways of breaking into cases: one can exhaustively enumerate all the different cases, in a flat list-like structure (C2). Or one can impose a hierarchical structure on the cases, with sister subcases being derived from the same mother case by adding extra information (C14). The first strategy makes for shallower proofs, with repetitive patterns of rule use (Assume–CTA); the latter makes for deeper proofs, with characteristic 'case opening' (Assume–Apply) and 'case closing' (Exhaust–Merge–Inspect) sequences. Deeper proofs actually require more steps, because the intermediate levels in the hierarchical structure of the proof are made explicit. Deeper proofs with subcase structure also require abstract situations to act as superordinate cases; hence, there will be more graphical abstraction in the proofs of subjects who generate proofs with this sort of nested structure.

As we stated in the introduction, our theoretical work has emphasised the idea that graphical systems possess a useful property—over-specificity—whereby certain classes of information must be specified. The property is seen as useful because inference with such specific representations can be very simple. We have also urged that real systems, like Hyperproof, do in fact allow abstractions to be expressed, and it is this that endows them with a usable level of expressive power.

From the case studies we have discussed, and from the empirical results which lie behind them, we can see that it is not enough for an inferential system simply to possess usable graphical abstractions. They must be *available* to the users of that system. DetHi subjects can exploit Hyperproof's graphical variables to create elegant proofs on indeterminate problems; DetLo subjects appear to lack the required competence with graphical variables, and so they attack indeterminate and determinate problems alike with roughly the same strategy.

The educational implications of this is are far from clear. Should all students be taught to use graphical reasoning methods, or should students be encouraged to follow their existing representational modality preferences? Much depends upon whether these prior cognitive styles are themselves responsive to educational intervention. Perhaps students should be encouraged to broaden their representational repertoires, before they encounter any formal logic teaching. For the time being, however, it seems that heterogeneous reasoning is bound to produce heterogeneous outcomes.

Acknowledgements

The support of the Economic and Social Research Council for HCRC is gratefully acknowledged. The work was supported by NATO Collaborative research grant 910954, and UK Joint Councils Initiative grant G9018050. The first author is supported by an EPSRC Advanced Fellowship. Special thanks to John Etchemendy, Tom Burke, and Mark Greaves.

References

Barwise, Jon, and John Etchemendy. 1994. *Hyperproof.* CSLI Lecture Notes Number 42. Stanford: CSLI Publications.

Cox, Richard, Keith Stenning, and Jon Oberlander. 1994. Graphical effects in learning logic: reasoning, representation and individual differences. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, 237–242.

Duran, Richard, Donald Powers, and Spencer Swinton. 1987. Construct Validity of the GRE Analytical Test: A Resource Document. Research Report

87-11. Princeton, New Jersey: Educational Testing Service, April 1987.

Stenning, Keith, and Jon Oberlander. 1991. Reasoning with Words, Pictures and Calculi: computation versus justification. In *Situation Theory and Its Applications, II*, ed. Jon Barwise, Jean Mark Gawron, Gordon Plotkin, and Syun Tutiya. 607–621. CSLI Lecture Notes Number 22. Stanford: CSLI Publications.

Stenning, Keith, and Jon Oberlander. 1995. A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive Science* 19:97–140. Also Research Report HCRC/RP-20, Human Communication Research Centre, University of Edinburgh, April 1992.

29

Austinian Pluralities

DICK OEHRLE

Plurality raises challenging questions—questions of language, questions of logic. The general study of grammatical number and quantity has a long tradition. The relatively recent rise of model-theoretic investigations of natural language interpretation has found a natural focus in the study of plural noun phrases, as witnessed by the insights of the work of Bennett, Scha, Link, and their many successors. This work has explored the range of plural *np* denotations, the contribution to distributive and collective interpretations of verbal information, the effect on these interpretations of adverbial modification, and the interaction of plurality with issues connected with the study of polyadic quantification and branching quantifiers.

An admirable survey of this direction of research is to be found in van der Does 1992 and van der Does 1993, where it is proposed that the range of plural readings described in the literature on plurality can be described by six 'lifts'—operations which map determiners of type $((et)((et)t))$ (the type, in the generalized quantifier tradition, of binary relations on sets of individuals) to determiners of type $((et)(((et)t)t))$ (the type of relations between sets and sets of sets). These lifts provide analyses of distributive, collective, and neutral readings, and do so in ways that take into account questions of relative scope. They take the following forms (where 'D' is a variable of type $((et)((et)t))$, 'X' and 'Z' are variables of type (et), '$\mathbf{Y}$' is a variable of type $((et)t)$, '$Pow(A)$' is the power set of A, '$\text{AT}(A)$' is the set of atoms of A):

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

$$\begin{array}{ll} D_1 & \lambda D \lambda X \lambda \mathbf{Y}.D(X)(\bigcup(\mathbf{Y} \cap \textsc{at}(X))) \\ N_2 & \lambda D \lambda X \lambda \mathbf{Y}.D(X)(\bigcup(\mathbf{Y} \cap Pow(X))) \\ N_3 & \lambda D \lambda X \lambda \mathbf{Y}.D(X)(X \cap \bigcup(\mathbf{Y})) \\ D_1^a & \lambda D \lambda X \lambda \mathbf{Y}.EZ \subseteq X[D(X)(Z) \wedge \textsc{at}(Z) \subseteq \mathbf{Y}] \\ C_2^a & \lambda D \lambda X \lambda \mathbf{Y}.EZ \subseteq X[D(X)(Z) \wedge Z \in \mathbf{Y}] \\ N_3^a & \lambda D \lambda X \lambda \mathbf{Y}.EZ \subseteq X[D(X)(Z) \wedge Z \subseteq \bigcup(\mathbf{Y})] \end{array}$$

For example, if the singular determiner D is 'three', the lift D_1 associates it with the plural determiner which combines with the interpretation of a noun such as 'boys' and the interpretation of a predicate such as 'smiled' to yield a proposition which is true iff a trio of boys is such that each one smiled. The other lifts characterize other readings.

One may accept the classification that Van der Does provides and nevertheless wonder why six lifts are appropriate and why just these six, rather than some other six that the language of higher-order type theory makes available. This paper is an attempt to provide a perspective on plurality which answers some of these questions.

29.1 Perspective

The perspective we offer on plurality is based on two simple ideas.

29.1.1 Austinian Propositions

First, we base our investigation on 'Austinian' propositions rather than 'Russellian' propositions (in the sense of Barwise and Etchemendy 1987). The most salient difference between these two conceptions—and the difference between them most relevant here—is that an Austinian proposition A involves both a situation s_A and a type of situation T_A, dependent on 'demonstrative conventions' and 'descriptive conventions', whereas a Russellian proposition lacks any contextually-determinable distinction among different possible situations. Barwise & Etchemendy introduce the useful notation $\{s; T\}$ to stand for 'that set-theoretic object completely determined by the situation s and [situation-type] T' [p. 124]. The possibility of representing *both* situation and situation-type is essential to the developments that follow.

29.1.2 Sums of Individuals; Sums of Situations

Second, we assume that *both* individuals and situations constitute domains structured by a *sum* operation. (Technically, we shall take each domain to be the carrier-set of a complete, but not atomic, join semi-lattice.) The correspondences between the classification of nouns as *count* / *mass* / *plural* and the classification of simple sentences as *telic* / *stative* / *iterative* according to what adverbials they are compatible with has become almost a commonplace in the linguistics literature (a commonplace based on the work of Verkuyl, Dowty, Bach, Krifka, Hinrichs, and others). Our focus here is not in the details of these structures, but the interaction of the two as a source of plural interpretations. Thus, we shall suppose that it makes sense to think of an Austinian proposition as being about a *structured set* of situations rather than a single situation, and that such a proposition may involve similarly-structured sets of individuals as the arguments of the relation involved.

29.1.3 Notation

If np_1 and np_2 are plural noun phrases, and v is a transitive verb, we represent the interpretation of each noun phrase as a join of individuals satisfying the interpretive constraints of its component parts and we represent the interpretation of the sentence as a join of situations satisfying the binary relation v' associated with v. We link these structures using a common index set I, together with indexed indeterminates. This yields the following data (where we assume that the relation '$\alpha \models \beta$' holds when α is a minimal model of β):

- $\bigsqcup_{i\in I}(s_i \models (v'(x_i^1)(y_i^2)))$
- $(\bigsqcup_{i\in I}(x_i^1)) \models np_1'$
- $(\bigsqcup_{i\in I}(y_i)^2) \models np_2'$

For example, the sentence *Four students bought three flutes* yields the data:

- $\bigsqcup_{i\in I}(s_i \models (\mathbf{bought}'(x_i)(y_i)))$
- $(\bigsqcup_{i\in I}(x_i)) \models (\mathbf{four\ students})'$
- $(\bigsqcup_{i\in I}(y_i)) \models (\mathbf{three\ flutes})'$

Reasoning about particular cases depends on properties of the relation v'. But it is already possible to identify different 'readings'—patterns of interpretation—with particular modes of indexation, since the index set I governs the correspondence between individuals and situations. And in fact, it is clear on reflection that intuitions about notions such as 'collective' in fact rest on access to this correspondence.

29.2 Development

29.2.1 Collective

We say an indexation is *collective* with respect to a given argument position np_k if for each $i \in I, x_i^k = \bigsqcup_{i\in I} x_i^k$ and adding, if we wish not to regard singular arguments as collective, the condition: x_i^k is itself a proper join. The interpretation of *Four students bought three flutes* can thus be collective with respect to either argument or both arguments together.

29.2.2 Distributive

We say an indexation is *distributive* with respect to a given argument position np_k if for each $i \in I, x_i^k$ is an atomic individual and $\forall i, j \in I, x_i^k \neq x_j^k$. One may also define other comparable notions here: each situation s_i might be such that x_i^k is atomic, without the requirement that $i \neq j \Rightarrow x_i^k \neq x_j^k$.

29.2.3 Branching

The 'branching' reading on which a sentence such as *Seven tourists visited six museums* is true if there is a set of tourists of cardinality 7 and a set of museums of cardinality 6 and each of the tourists visited at least one museum and each of the museums was visited by at least one of the tourists is simply a special case of the schema given above. For example, if each of the visits consists of one tourist visiting one museum, the schema given above requires that the sum of the visiting tourists be a model of *seven tourists* and that the sum of the visited museums be a model of *six museums*.

29.2.4 Asymmetries of Indexation

In some cases, it is useful to ask whether the index set is determined by the structure of one or the other of the arguments or by the structure of the situation associated with the proposition. In the *cumulative* readings of Scha 1984, exemplified by such sentences as *5000 Dutch companies own 9500 workstations*, we may regard the index set I as determined by the structure of the set of Dutch companies. Similarly, it may be fruitful to adopt the opposite perspective with regard to examples of the sort *Four thousand ships passed the lock*, analyzed by Krifka 1990: that is, we may regard the index set in these cases as determined by the cardinality of the set of lock-passings.

29.2.5 Referential Relations

It is interesting to consider the interpretation of referential relations from the point of view of our basic schema. There are in fact two sites in which referential relations may be constrained: either within the relational schema associated with s_i or across the argument representations $\bigsqcup_{i \in I} x_i$ and $\bigsqcup_{i \in I} y_i$.

A simple representation of the interpretation of English (plural) reflexives and reciprocals as objects of a transitive verb in a sentence of the form $np_1\ v\ np_{ref/rec}$ may be formulated as follows:

- **reflexives:**
  - $\bigsqcup_{i \in I} (s_i \models (v'(x_i)(y_i)) \wedge x_i = y_i)$
  - $(\bigsqcup_{i \in I}(x_i)) \models np'_1$

- $(\bigsqcup_{i \in I}(y_i)) = \bigsqcup_{i \in I}(x_i)$

- **reciprocals:**
  - $\bigsqcup_{i \in I}(s_i \models (v'(x_i)(y_i)) \wedge x_i \neq y_i)$
  - $(\bigsqcup_{i \in I}(x_i)) \models np'_1$
  - $(\bigsqcup_{i \in I}(y_i)) = \bigsqcup_{i \in I}(x_i)$

These two schemas agree in equating the interpretation of the subject and object arguments at the global level. In addition, they both impose a condition on the relational schema supported by each s_i: reflexives require that the relation hold of identical pairs $< x_i, x_i >$; reciprocals require that the relation hold of distinct pairs $< x_i, y_i >$.

These formulations have interesting consequences. For example, on the reflexive interpretation, a sentence such as *Smith and Jones nominated themselves* may be true (a) if Smith nominated herself and Jones nominated herself or (b) if Smith and Jones collectively nominated Smith and Jones together. But it correctly rules out the possibility of using *Smith and Jones blamed themselves* of a situation in which Smith blamed Jones and Jones blamed Smith. These simple consequences are difficult to formulate in the absence of a theory which has access both to the structure of the relevant relation-supporting situations and the structure of the relation's arguments.

The account of reciprocals also has points of interest. Note first that reciprocals cannot be bound by singular noun phrases. This is a direct consequence on the formulation here of the obvious unsatisfiability of the inequality $x \neq x$ (from the relational schema) and the equality $x = x$ (from the identification of the two arguments). Second, the account here offers insight into the relation between reciprocal interpretations in particular and plural interpretations in general. It is a direct consequence of this connection that the account of reciprocals formulated here is compatible not only with interpretations of the form $(A \times A) \setminus \Delta$—the cross-product of the argument interpretation A with itself minus the diagonal Δ of $A \times A$—but also with much weaker relations among the members of the argument interpretation A, as studied by Langendoen 1978. On standard (Russellian) accounts, the range of reciprocal interpretations poses a puzzling question. Our formulation treats almost all of these special 'readings' simply as special cases of the single general and simple schema above (almost all, because cases like *The plates are stacked on top of each other* still resist accommodation).

29.2.6 Different / Same

The contribution of the 'quantificational adjective' *different* to the interpretation of sentences such as *Smith and Jones read different books* or *Dif-*

ferent students walked and talked can be treated straightforwardly from the present perspective. This is of interest because it offers a simple alternative to account based on polyadic quantification.

If *different books* corresponds to the k-th argument of a relation R, we need to impose two interpretive conditions:

- $\bigsqcup_{i\in I} x_i^k$ is a proper join of atoms.
- $\forall i, j \in I, x_i^k \neq x_j^k$

To see the consequences of these conditions, the sentence *Smith and Jones read different books* determines the following schema:

- $\bigsqcup_{i\in I}(s_i \models (\mathbf{read}'(x_i)(y_i)))$
- $(\bigsqcup_{i\in I}(x_i)) \models (\mathbf{smith}' \sqcup \mathbf{jones}')$
- $(\bigsqcup_{i\in I}(y_i)) \models (\mathbf{books})', \forall i, j \in I, y_i \neq y_j$

Now note that any of the readings s_i must be a reading by either Smith or by Jones; in addition, there are readings by each; finally, the inequality imposed (locally!) by *different* demands that each reading involve a different book. The point of essential interest here is that the locally imposed constraints have global consequences through their effect on the properties of indexation. The interpretation of comparable quantificational adjectives such as *same* can be treated along similar lines.

29.3 Future Directions

The attraction of the point of view elaborated above is that it provides a unifying perspective on a broad range of phenomena related to the interpretation of plurality: that is, different 'readings' arise as special cases of a single basic schema. The inherent properties of this basic point of view do not cover, however, the entire range of phenomena involving questions of plurality. Whether the unifying perspective proposed here is consistent with adequate accounts of this broader range of issues depends both on future development of this line of research and on insight into the nature of the empirical issues involved. We comment below on three areas of interest—coordination, quantification, and the pragmatics of plural interpretation.

29.3.1 Coordination

A problem for theories which characterize the distinction between distributive and collective interpretations as a distinction between 'plural noun phrases' and 'collective noun phrases' is that a single occurrence of a noun phrase can apparently be interpreted as 'distributive' with regard to one predicate and 'collective' with regard to another. As an example, consider sentences such as

Three students walked in and pooled their resources.

Using the notation of §1.3, we have, as a first approximation:

- $\bigsqcup_{i \in I}(s_i \models (\textbf{walked in and pooled their resources}'(x_i)))$
- $(\bigsqcup_{i \in I}(x_i)) \models (\textbf{three students})'$

At this level of analysis, it is obviously impossible to reconcile this representation with the possibility of relating the three students distributively to three entrances and relating them collectively to a single pooling of resources. But at a more fine-grained level of scale, nothing prevents this. Indeed, suppose we represent the tensed *vp*'s *walked in* and *pooled their resources* and their conjunction as follows (using '$\sqcup$' for disjoint union):

$$\bigsqcup_{i \in I}(s_i \models (\textbf{walked in}'(x_i)))$$
$$\bigsqcup_{j \in J}(s_j \models (\textbf{pooled their resources}'(x_j)))$$
$$\bigsqcup_{k \in (I \sqcup J)}(s_k \models (\textbf{walked in and pooled their resources}'(x_k)))$$

The corresponding condition on the subject *np* is given below:

$$(\bigsqcup_{k \in (I \sqcup J)}(x_k) \models (\textbf{three students})'$$

If the cardinality of the index set I is 3, and each $i \in I$ is associated with a different student, and the cardinality of J is 1, and $j \in J$ is associated with the sum of the three students indexed by $i \in I$, we have a representation of the desired interpretation.

One observation consistent with this approach (due to Arthur Merin) is that *vp* conjunction is not always idempotent: *The bell rang and rang* is not equivalent with *The bell rang*, since the latter is consistent with a single toll of the bell, but the former is not. If we take *vp* conjunction to involve Austinian propositions (or schemata abstracted from them) of the form $\{s_1 \sqcup s_2; \ldots\}$, where the situation involved is a disjoint union of two sub-situations, then there is a straightforward explanation of this interesting property. This phenomenon deserves much more extensive investigation, particularly in view of the existence of languages in which a single morpheme on a verbal stem is interpretable as plurality of a specified argument or plurality of events.

On the theoretical side, the analysis sketched in a rudimentary way above of the interaction of plural arguments and conjoined predicates needs to be examined much more extensively, particularly with regard to the interaction of conjoined predicates with the range of plural readings enumerated above in §2, and with regard to interpretive constraints imposed by such adverbs as *respectively*.

29.3.2 Quantification

Along with the interpretations discussed earlier, the sentence *Four boys bought three flutes* has quantificational interpretations, such as the interpretation on which each of four boys bought three flutes. These interpretations cannot be expressed using the schema for 2-place predicates introduced in §1.3, because the sum of the flutes involved may exceed three. It is possible, however, to represent such readings in a way consistent with the general approach advocated here:

$$\begin{aligned}&\textstyle\bigsqcup_{\imath \in I}(s_\imath \models \lambda z.(\bigsqcup_{\jmath \in J_\imath}(s_\jmath \models \mathbf{bought}'(z)(y_\jmath)))(x_\imath))\\ &\qquad (\textstyle\bigsqcup_{\jmath \in J_\imath}(y_\jmath)) \models (\mathbf{3\ flutes})'\\ &(\textstyle\bigsqcup_{\imath \in I}(x_\imath) \models (\mathbf{4\ students})'\end{aligned}$$

Additionally, it would seem, it is necessary to suppose that the index function mapping $\imath \mapsto s_\imath$ is injective.

An attractive possibility, but one which must be left to future work, is that the necessary adjustments to deal with quantification arise in the context of a general type theory, as special cases of general type-shifting rules. For example, using e to represent the type of individuals and s for the type of situations, we might take two-place relations to belong to a family of types: in particular, both $(e \cdot e) \rightarrow s$ and its curried variants of the form $e \rightarrow (e \rightarrow s)$. Non-quantificational *np*'s have type e and are liftable to the type of quantificational *np*'s: $(e \rightarrow s) \rightarrow s$. (For a categorial grammar allowing these type-shifting rules which straightforwardly accommodates quantificational scope ambiguities in simple clauses, see Oehrle 1995. One question for future work is whether or not appropriate interpretive properties can be assigned to these types according to general principles—that is, along the lines of the Curry-Howard correspondence. A second question is whether an analysis of simple quantificational interpretations of the kind given in the representation above can lead to deeper insight into the role of Conservativity in natural language quantification. We leave these issues open here.

29.3.3 Pragmatics of Plural Interpretation

An important question for all theories of plural interpretation is why so many apparently diverse interpretations are available, while particular utterances involving plural interpretation often seem to give rise to no difficulty in interpretation. In some cases, intonational structure and prosodic phrasing might provide disambiguating cues. But this still leaves many possible models for simple sentences. A plausible approach delegates this issue to the interplay between what is said and what is conveyed, along Gricean lines. Craig 1994 discusses this problem and explores how a Gricean approach might work.

29.4 Summary

The research reported here is still in the exploratory phase. It is based on three simple ideas: first, adopting an Austinian account of propositions; second, considering both situations and individuals as structured by comparable join-operations; third, investigating the relations between these structures as a resource giving rise to interpretive fine-structure in the analysis of plurality, quantification, and the issues of referential relations that these phenomena give rise to. On the basis of this preliminary work, this combination of simple ideas seems worth pursuing further.

References

Barwise, Jon, and John Etchemendy. 1987. *The Liar: An Essay on Truth and Circularity.* Oxford: Oxford University Press.

Craig, Thomas B. 1994. The Pragmatics of Scalar Quantifier Scope Interaction. Seminar paper, Department of Linguistics, University of Arizona.

Krifka, Manfred. 1990. Four Thousand Ships Passed through the Lock: Object-Induced Measure Functions on Events. *Linguistics and Philosophy* 13(5):487–520.

Langendoen, D. Terence. 1978. The Logic of Reciprocity. *Linguistics Inquiry* 9(2):177–197.

Oehrle, Richard T. 1995. Term-labeled Categorial Type Systems. *Linguistics and Philosophy* 17(6):633–678.

Scha, Remko J. H. 1984. Distributive, Collective and Cumulative Quantification. In *Truth, Interpretation and Information: Selected Papers from the Third Amsterdam Colloquium*, ed. Jeroen Groenendijk, Theo M. V Janssen, and Martin Stokhof, 131–158. Dordrecht. Foris.

van der Does, Jaap. 1992. *Applied Quantifier Logics.* Doctoral dissertation, Universiteit van Amsterdam.

van der Does, Jaap. 1993. Sums and Quantifiers. *Linguistics and Philosophy* 16(5):509–550.

30

Interfacing Situations

John Perry and Elizabeth Macken

Introduction

This paper is the first in a series of two, in which we (i) explore some aspects of *heterogeneous systems of representation and communication*; (ii) show how American Sign Language (ASL) exhibits some of those features; (iii) draw some morals for the design of interfaces. This paper explores (i) at some length and ends with a brief look at (ii).

Heterogenous systems of representation and communication are systems that combine representations whose meanings work on different principles, such as pictures and words. (We will try to reserve the word "language" for natural languages, like English and American Sign Language (ASL), and not use it for just any system of structured representations.)

This paper reflects work that we have been doing in collaboration with Cathy Haas of the Archimedes Project at CSLI and Bill Stokoe of Galludet University, having to do with *richly grounded meaning* in ASL.

Richly grounded meaning or RGM is a generalization of what Peirce called "iconicity"; the symbol and what it symbolizes are naturally rather than arbitrarily connected.[1] The key word here is "arbitrary"; probably most RGM symbols are conventional in the sense developed by David Lewis in *Convention* (1969), but there is a natural connection between the symbol and what it symbolizes. The traditional word instead of "natural" might be "resemblance". We emphasize that what is in question is something psychological; a robust cognitive correspondence between properties of a symbol (which must have enough interesting properties to ground such a relation, hence "richly grounded") and properties of that which is symbolized. Resemblance is too restrictive. There are, we think, various logical properties

[1] The concept of RGM is described more fully in (Macken et al. 1993).

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

that correlate with, contribute to, and may even be necessary conditions of such natural correspondences, but whether a system that meets the logical requirements actually works to give an intuitive system is always a matter of psychological fact. A correspondence that is cognitively robust for one person may not be for another, due to different experiences stemming from difference in training, culture, and the like.

Heterogeneous communication systems combine arbitrary-conventional symbols with richly grounded ones. Maps, charts, Hyperproof[2], graphical user interfaces and American Sign Language are all, we think, heterogeneous communication systems.[3]

As the examples of Hyperprooof and sylized documents suggest, there has been a lot of thinking about heterogeneous languages going on by people attending STASS conferences, and we will draw particularly on work from STASS II by Barwise and Etchemendy (1990b, 1990a) and Stenning and Oberlander (1991.)

In all of the cases, symbols with arbitrary-conventional meaning (ACM) are combined with symbols with RGM. For example, the upper part of a Hyperproof screen has a pictorial diagram of a blocks world, while the bottom part gives information about the same situation in the predicate calculus. The top part may contain labels, ACM symbols that allow the integration of information. In a map, symbols with ACM (like "Lincoln" and "Omaha" are placed in a 2-dimensional representational system (or are used to label dots in such a system) in which distance and direction are used to represent distance and direction.

We do the following:

- Discuss criteria that have been offered for what makes representations diagram-like or picture-like. We will look at Barwise and Etchemendy, Larkin and Simon, and Stenning and Oberlander.
- On the basis of ideas and examples gathered from these authors, we will provide a list of criteria which allow us to distinguish among text-like, chart-like, diagram-like map-like and picture-like representations.
- We will then show how ASL incorporates text-like, chart-like and diagram-like systems of representation.

Before plunging into this, however, we want to put it in a larger perspective, by briefly describing the Archimedes Project at CSLI — and the present state of the computer and communications industry— and how we

[2]Hyperproof is system for teaching basic elements of reasoning, developed by Jon Barwise and John Etchemendy (1994).

[3]We also think that stylized documents of the sort discussed by Devlin and Rosenberg fit into this category, although we won't discuss them tonight.

see these as motivating the study of heterogeneous systems of representation and communication.

30.1 The Archimedes Project and Heterogeneous Communication Systems

The Archimedes Project at CSLI has to do with the accessibility of information to people with disabilities. It turns out that many themes from the situation-theoretical perspective are quite relevant to and we think helpful in thinking about this, in particular the distinction between information and particular ways of presenting information.

We use the term "disabled" for individuals with some condition or injury that prevents them from picking up information or initiating action in one or more of the standard ways. "Handicapped" means that one cannot pick up information that most people around you can, or cannot do the things most people around you can. Disability is often a contributing factor to being handicapped, but it is neither a necessary nor a sufficient condition. We are handicapped when we travel in Japan, because there is loads of information of which we cannot take advantage, not knowing Japanese. An individual with a disability that requires her to use a wheelchair may have no difficulty getting information from a computer screen and inaugurating actions with a keyboard and mouse; she is not handicapped by her disability in this activity.

A central idea of the Archimedes Project is that people with disabilities are quite unnecessarily handicapped by systems that make information accessible only in one form, suited for a particular mix of perceptual abilities, or requiring a specific motor function to inaugurate action.

A recent example of this is "The GUI Problem" for blind computer users. Blind people work well with computers with DOS interfaces. Screen readers automatically convert the ASCII code to voice. But Graphical User Interfaces or GUIs have by and large been a disaster for them. Screen readers fail when they hit windows, pull-down menus, icons, and the like. "Road-kill on the information highway," as the blind scientist Larry Scadden said recently about the his adventures on the world-wide web.

The concept of *heterogeneous systems of communication* developed as we studied this problem. Our idea is that if we knew more about why such systems in general, and GUIs in particular, were popular, and seemed to help people work with and communicate information more efficiently, we would be better able to understand what sort accessible alternatives might be envisaged that would provide the same functionality.

Now it turns out that there is another example of a heterogeneous language that is in fact a great boon for people with disabilities, namely American Sign Language (ASL) which is used by many deaf people to communicate. ASL makes a particularly interesting object of study because it is a natural language, with all of the expressive power and subtlety of English or any other language. But because it is a language of gesture rather than sound, it provides many more opportunities for richly grounded meaning. Skilled ASL users employ the stage-like possibilities of signing in clever ways, and have various techniques for integrating information.

So our interest is really based on two hypotheses. The first is that heterogeneous systems of communication arise whenever the possibility of richly grounded meaning is available. Neither spoken language by itself nor written language confined to text provide many opportunities, as the signal is rather thin. But when people are face-to-face and can use gestures, bits of paper, napkin, blackboards or whatever, they do so.

Before turning to our main topic, the factors that make things more and less picture-like, let us say a little more about RGM generally.

It is important to distinguish RGM in elements and *higher-order* RGM. By an element, we mean representations of particular objects, such as proper names in text, dots for cities on a map, and the parts of a picture that represent specific things. An element has RGM if it resembles (i.e., cognitively corresponds to) the object that it stands for. Many of the signs of ASL, and of other signing systems, strike people as having this property. However, there is an extensive literature arguing that this property does not play a very important role in ASL (Frishburg 1975, Klima and Bellugi 1979). Although many signs may begin as iconic, they tend to become sylized. They lose the psychological features we associate with RGM signs, readily inferable meaning (RIM) and easily remembered meaning (ERM).[4]

We will call a system of representations "iconic" if it has lexical RGM. We distinguish three levels of iconicity. Pictures are fully iconic. The link between representations and what they represent is based on resemblance or some other cognitively robust connection. Maps are partly iconic. Many of the symbols on a map have no resemblance to what they represent, but many do, particularly depictions of rivers and highways. Ordinary text is non-iconic. With a few exceptions, it is only arbitrary conventions which connect the basic symbols with what they represent.

By higher-order RGM, we have two things in mind. First, the pervasive use in ASL of what we call Internally Modifiable Meaning (IMM). Even if

[4]Without in any way doubting the validity of this literature, we can also observe that the fight to rid ASL of its image as a system of icons coincided with the fight to have it recognized as a full-fledged language. From the time of Stokoe's pioneering work in the 60's through to the present, the emphasis in ASL studies has been on the similarities between ASL and spoken languages.

a sign has little or no RGM, there may be a system of modifications to it that do.

The ASL sign for *improve* is an example of this. The basic sign involves moving the right hand, palm facing back, first to the wrist and then near the crook of the extended left arm. This sign is basically ACM—although there is a certain naturalness in having an upwards movement as part of the sign for improvement. The important point is how modifications, which in English would be handled with adverbs, are handled. The amount of improvement may be signed by the relative distance moved on the forearm. This modification is readily inferable—anyone who knows the sign for improve will easily grasp what the modified signs mean.

The second thing we have in mind by higher-order RGM is the presence in ASL of a mode of signing that is closer to drawing pictures or diagrams or maps than to constructing phrases or sentences, and this is what we will explore tonight.

30.2 What makes representations picture-like?

We turn now to the question of what makes a visually inspectable representation picture-like, map-like, diagram-like, chart-like or text-like.

As a paradigm, think of the Tarski's World or Hyperproof display. The top of the screen is picture-like (or tends towards being picture-like), and the bottom is text-like. The example on which we focus here is taken from Example 5 in Barwise and Etchemendy's "Visual Information and Valid Reasoning" (1990b).

Here are two representations of the same situation:

$$\begin{array}{ccccc} \text{A} & \times & \text{C} & \text{B} & \text{D} \\ \text{—} & \text{—} & \text{—} & \text{—} & \text{—} \end{array}$$

FIGURE 1 A diagram of five chairs and their occupants

a $\neq$ b, a $\neq$ c, etc.
a is a chair, b is a chair, c is a chair, etc.
a is to the left of b, b is to the left of c, etc.
A is a person, B is a person, etc.
A sits in a, B sits in b, etc.
b is not to be occupied.

FIGURE 2 *A textual description of five chairs and their occupants*

Figure 1 depicts and Figure 2 describes a situation in which five chairs are arranged in a row and a person A occupies the leftmost chair, no one is allowed to occupy the second from the left, and C, B and D occupy the next three. The people are indicated by large letters in both figures; the chairs are indicated by lines in the first and small letters in the second.

What makes the first figure more picture-like, the second more text-like? We will begin our investigation by examining a list of criteria offered by Barwise and Etchemendy.

30.2.1 Barwise and Etchemendy's Criteria

In (1990b), Barwise and Etchemendy list six ways in which diagrammatic representation differs from linguistic representation: the former exhibit closure under constraints, conjunctive rather than disjunctive information, and homomorphic representation. They support symmetry arguments and perceptual inference.

The point with respect to symmetry arguments is that such arguments are often involved in reasoning with diagrams (for example the reasoning problem connected with Example 5). This point about reasoning with diagrams is not presented as either a necessary or a sufficient condition for something being a diagrammatic representation, so we are going to set it aside. The point about perceptual inference we will defer until later.

(We should emphasize that we are not indulging in the old-fashioned philosophical exercise of searching for the essence of pictures or diagrams or RGM. We are engaged in the new-fangled cognitive science exercise of looking for contributing factors to differences that we intuitively feel and exploit, that will lead to better and more useful classifications of the phenomena in which those differences are found, and may support increasingly detailed empirical and mathematical studies of the phenomena.)

So, as we were saying, what is the essence of pictures?

Homomorphism is at best a necessary condition. If we consider Figure 1 and Figure 2 we have homomorphism in both cases. We will make the point by showing the correspondence between the representing facts and the represented facts. We will call the lines in Figure 1, "1","2","3","4" and "5"; we will designate the people and chairs with large and small letters, respectively.

For the pictorial representation:

Not same line, 1,2	Not same chair, a, b
etc.	etc.
Is line, 1	Is chair, a
etc.	etc.
Is big letter, "A"	Is person, A
etc.	etc.
Is-left-of, "a","b"	Is-left-of, a,b
etc.	etc.
Is-above, "A", 1	Sits on, A,a
etc.	etc.
Has × above, 2	Is not to be occupied, b

FIGURE 3 Homomorphism from diagram to chair situation

For the linguistic representation:

Have tokens flanking "≠", "a","b"	Not same chair, a,b
etc.	etc.
Has token before token of "is a chair", "a"	Is chair, a
etc.	etc.
Has token followed by token of "is a person", "A"	Is person, A
etc.	etc.
Have tokens flanking token of "is to the left of" "a","b"	Is-left-of, a,b
etc.	etc.
Have tokens flanking "sits in", "A", "a"	Sits on, A,a
etc.	etc.
Has token followed by token of "is not to be occupied, "b"	Is not to be occupied, b

FIGURE 4 Homomorphism from description to chair situation

Now Wittgenstein, noticing something like the sort of homomorphism we just presented, advanced the idea that sentences were pictures (1922; see also Etchemendy 1976). He might be right at a suitably deep level, but at the level at which we are operating, we draw the conclusion that a homomorphism between the representing and the represented is not enough to make the representation picture-like.

In the case of real pictures, it is not so clear that there is a perfect homomorphism. In a picture that uses perspective, one element being above another can signify that one thing is behind another or that one thing is above another. This may simply show that we have not chosen the representing relations carefully enough to find the homomorphism. We will assume that homomorphism is a good approximation of a necessary condition for being picture-like or diagram-like.

Now let's look at "closure under constraints". As Barwise and Etchemendy note, diagrams are physical situations and so they obey their own set of constraints. They say,

> By using a representational scheme appropriately, so that the constraints on the diagrams have a good match with the constraints on the described situation, the diagram can generate a lot of information that the user never need infer. Rather, the user can simply read off facts from the diagram as needed. This situation is in stark contrast to sentential inference, where even the most trivial consequence needs to be inferred explicitly (1990b).

As we understand it, the property in question is more fully describable as,

> Constraints on the facts in a representation R that represent facts about a relation Q are such that IF Q-facts $f_1 \ldots f_n$ are explicitly represented in R, and $f_1 \ldots f_n$ guarantee f-fact f_{n+1}, THEN R will explicitly represent f_{n+1}.

Here is an example. Let these three blocks be our situation:

FIGURE 5 Structure of facts of three-blocks situation

Let our picture-like representation be based on the idea that the representation will be a row of letters on a line from left to right, so that a letter being to the left of another represents the fact that the represented blocks are in the left of relation.

A B C

FIGURE 6 Diagram-like representation of three-blocks situation

Let our language-like representation be a sequence of sentences of the form "X is to the left of Y". If a sequence of letters X,Y flanks the "is to the left of" predicate, that represents that the block X stands for is to the left of the block Y stands for.

Now if we put in our diagram-like representation a representation to the effect that the box is to the left of the diamond, and one to the effect that the diamond is to the left of the triangle, we will have eo ipso put one in to the effect that the box is to the left of the triangle.

But, if we write the sentence "B is the left of D" and the sentence "D is to the left of T" we will not have thereby, eo ipso, written the sentence "B is to the left of T".

B is to the left of D
D is to the left of T

FIGURE 7 Linguistic representation of three-blocks situation

So our diagram-like representation is closed under constraints, and our language-like one is not.

Why is this so? In Figure 6 the transitivity of the "is to the left of in a row" relation among tokens of letters matches the transitivity of being *to the left of in a sequence of blocks. But as Figure 7 shows, the relation* of having letters that flank the words "is to the left of" is not transitive. The relation holds between "B" and "D" and between "D" and "T" but not between "B" and "T".

Closure under constraints is a real difference between a diagram and a typical representation that is not diagram-like. But it is not a logically sufficient condition for being diagram-like. One can imagine a magic slate that always automatically produced the closing representation—i.e., would just write "A is to the left of C" when someone had written on it, "A is to the left of B" and "B is to the left of C". That would not be a picture-like representation.

(Approximate) homomorphism and closure under constraints arise when (but perhaps not only when) we have *systematic, constrained and localized representation.* This requires that three conditions are met. First, a whole system of relations is systematically interpreted as representing another system of representations, rather than the interpretations being assigned piecemeal. Second, the representing relationships obey the constraints that correspond to those obeyed by the represented relationships. Third, there is only one token for each individual object.

Consider a diagram one might draw to show someone how to lay out a croquet court. A great many croquet courts of different sizes and oriented in different directions might satisfy the diagram. It is the relative distances and relative directions that count. For each court that satisfies the diagram there will be a homomorpism between distances between wicket symbols on the map and distances on the court, and between orientation on the diagram and directions. The homomorphism is not fixed piecemeal; once it is fixed that one distance on the diagram represents a certain distance in the world, all the other interpretations are fixed, and similarly with directions.

One could have systematic interpretation in a text-like representation; the distance relations might all by expressed by inter-related linguistic formulae, such as "being n meters", "being n+1 meters" etc. But such a linguistic relation would not be closed under constraints.

Note that the oddity or unnaturalness of our text-like homomorphisms comes, at least in part, from the fact that we allow more that one token for a given object in a given representation.[5] Our diagram of a croquet court, however, meets what we call "the unique token requirement". There is one and only one representation for each wicket. All the representing facts about that wicket—the facts that represent its distance from other wickets, its direction, and any other facts about it that are represented—will involve that one representation.

Multiple-token representation is ubiquitous in language, of course. It has the effect of destroying the constraints that guarantee closure. Returning to the example involving Figure 5, if we had allowed ourselves to use two tokens of "A" in our representation of the row of shapes in Figure 6, then we could have had a representation that explicitly represented A being to the left of B, and B being to the left of C, without having an explicit representation of A being to the left of C.

Finally, Barwise and Etchemendy say that diagrammatic representations are conjunctive rather than disjunctive. This should not be taken to mean that a particular representing fact cannot represent a range of alternatives. There are many actual croquet courts, facing different directions and with different distances between the wickets, that satisfy the diagram we are imagining. The point is rather that the effect of adding a new representing fact to a picture or diagram-like representation is to conjoin a fact to what is already represented, not provide an alternative. This is a consequence of systematic, constrained and localized representation. One creates new representations by placing new representations for individual objects onto the diagram. The new representation cannot represent an alternative for one of the wickets already represented, by the unique token requirement.

30.2.2 Larkin and Simon

This property of unique token representation is related to the use of locations for grouping information, that Larkin and Simon emphasize (1987). They provide three reasons why diagrams can be superior to verbal descriptions for solving problems.

- Diagrams can group together all information that is used together, thus avoiding much searching for the elements needed to make a problem-solving inference.
- Diagrams typically use location to group information about a single element, avoiding the need to match symbolic labels.

[5] Along these lines, John Etchemendy has called the diagrammatic part of the Hyperproof display "token based".

- Diagrams automatically support a large number of perceptual inferences, which are extremely easy for humans.

The first two reasons emphasize the way diagrams use location to group information about a single object. This is lost when one uses the system of types and token. Many different tokens of the same type designating the same individual object may be scattered around a document, so that the information the document contains about that individual is not localized. It is a feature of perception that the perceptually accessible information about an individual is centered on that part of the perception that we think of as being *of* the individual. Monadic or intrinsic facts about the individual will be picked up by inspection of the individual, and relational facts will involve of part of the scene that involves the individual. This sort of localization makes looking at a diagram or picture *like* looking at the things themselves, and permits the inferential abilities of the perceptual system to be exercised on the diagram or picture.

30.2.3 Stenning and Oberlander

There are two ways that one could end up with two representations of the same object. One could have two tokens of a type, both of which designate the object. This is ruled out by the unique-token property. But one could also have two types assigned to the same object, like "Tully" and "Cicero". Then a representation could satisfy the unique token requirement, but still have multiple representations, and multiple loci of information, about a single object. In "Words, Pictures and Calculi," Stenning and Oberlander point out that it is a feature of graphical representations to not allow this.

Stenning and Oberlander find the difference between graphical representations and textual representations in a property they call specificity. We suggest that there are several aspects to specificity that are worth distinguishing. The major division is between *determinateness* and *regimentation.* Determinateness we further divide into two kinds, *issue determinateness* and *Berkeley determinateness.*

Determinateness. The basic idea of issue determinateness is that if a representation raises an issue, it settles it. Let our representation be the following two sentences:

- Madeline is charming. David works at SRI.

The representational resources of this representation allow us to raise two further issues: Is David charming? Does Madeline work at SRI? But the representation does not settle them.

We will say that an issue, in the situation-theoretical sense of a relation and a suitable sequence of arguments, is available from a representation if the representation contains items that stand for the relation and each of the arguments. Issue determinateness means that all available issues are settled by the representation—that is, that it explicitly represents that the answer for the issue is yes or no (polarity 1 or 0).

This property requires more of a picture or diagram than the property of closure under constraints that Barwise and Etchemendy mentioned. Suppose we have a representation of the fact that A is larger than B and a representation of the fact that C is below D. The closure condition does not require that we have representations that tell us whether or not A is below B, or C is larger than D, but this property does. However, it does seem that systematic, constrained and localized systems of representation meet this condition. In such a system, an element represents things about the object it designates in virtue of having various properties and standing in various relations. Each of the other elements will have properties of the same kind and stand in relationships of the same kind. So issues that are settled for one object, will be settled for all. For example, when one puts a dot representing Omaha on a map, making issues about Omaha available in our sense, that dot will be a certain distance from all the other dots. Putting the dot on the map, which makes the issues available, also settles them.

The second notion of determinateness is suggested by Stenning and Oberlander's citation of Berkeley, so we call it "Berkeley determinateness". What impressed Berkeley was the fact that you couldn't draw a picture of a triangle or have a mental image of a triangle that wasn't a picture or mental image of some definite type of triangle, scalene, isosceles, right angle, etc.

To state what Berkeley determinateness entails, we need the determinable-determinate distinction. This is exemplified by color and red, or height and 5'3", or shape and 3/4/5 right triangle or weight and 180 pounds. Any object that has a determinable property (shape, color, size) has some determinate value of it. But it is not the case in general that a representation that represents an object as having a determinable property represents the object as having some determinate value of it. If we say, "David has an interesting shape," I imply that he has a shape, but I don't say exactly what his shape is.

It is a property a representational system might have, relative to some category of properties, that if it represents an object as having a determinable property then it represents that object as having some determinate value of it. This is what we call Berkeley determinateness with respect to that category of properties.

However, it is not generally the case with pictures that they are Berkeley determinate with respect to the visually detectible properties they depict. An artist need not decide whether she is painting a picture of tall people standing in front of a tall tree or short people standing in front of a short tree. She represents the people and the tree as having height (and arguably, weight) but not specific heights and weights. They are represented as having a three-dimensional shape, but not as having specific one. But if she included a scale, and thus provided a systematically homomorphic representation, the picture would have this property.

The fact is that there are some correspondences that are so natural, that it is difficult to imagine alternatives. If you are going to use closed figures to represent shapes, what shape should a triange represent? It seems that the natural answer is, triangles. What should an isosoles triangle represent? An isosoles triangle. The representation will have a determinate value of each of its determinable properties, and so if it represents exactly the same properties it has, the representation will be Berkeley determinate.

But this doesn't generalize. If we are going to use objects with size to represent sizes, what size should a one inch figure represent? The answer is not so obvious. We are used to representing one size with another, and one distance with another, but not to representing one color with another, or one shape with another.

Given a systematic, constrained and localized system of representation, we need to fix rather than merely constrain the homormorphisms between representing and represented relations to get Berkeley specificity. With our croquet court diagram we could do this by adding a scale and a north arrow.

Our conclusion then is that systems that are systematic, constrained and localized need not be Berkeley determinate. That is an additional property. Some systems may have it because there is a very natural built in "homomorphism-fixer" that we assume at least as a default: red represents red, yellow represents yellow, etc. Other systems may have it because devices like a scale or a north arrow fix the homomorpism.

The fact that some representing properties such as colors and shapes seem to have a built in homomorphism-fixer can create problems for designers of representational systems. Lingraphica is a system designed by Dick Steele for the use of global aphasics, who have lost the ability through brain injury to remember the meanings of words. Steele designed an iconic, MacIntosh-based system for communicating with aphasics. He concentrated for a while on recipes, which he found that his patients could follow unassisted, by figuring out the meaning of the icons.

He had trouble coming up with an icon for "stir". The natural way to do it is with a dynamic icon showing a bowl of stuff being stirred by a spoon, say. But how to make this an icon for "stir" and not "stir with a spoon"?

Regimentation Stenning and Oberlander give this list of representation systems, to indicate some points along a dimension they call "regimentation":

> ...a quantified abstraction; a disorderly text; an orderly text;...an alphabetized table of intercity distances; the same table with cities ordered by longitude in the column labels and latitude in the row labels; and finally a map.

Stenning and Oberlander connect this dimension with the number of ways there are for making a representation true, and this connects regimentation with determinateness. We'll bypass the issues here, and note that at least one aspect of regimentation connects with localization and the unique token requirement.

In an ordinary piece of text, there are no limitations on the number of different tokens that might stand for a given individual object, nor any restriction on where they might occur. Still, a good orderly presentation will exhibit some localization of information. Consider for example the CSLI brochure that is given to prospective Industrial Affiliates. There is a paragraph headed by the name of each researcher, which contains certain vital facts about them. A reader might turn to this page to know who, for example, John Etchemendy is. But tokens of the name "John Etchemendy" might also occur elsewhere in the brochure; not all the information about Etchemendy is localized.

Stenning and Oberlander compare such more or less orderly texts with a mileage chart and a map. In a mileage chart, there are two tokens of each name, one labelling a column, one labelling a row. Finally, with a map, we have one token per object represented.

This suggests a dimension defined by the constraints on the number and location of tokens of names, with the unique token constraint at one end and the total lack of constraints found in disorderly text at the other. We will use just a rough and ready classification: no constraints, some constraints, and the unique token requirement.

30.3 From Texts to Pictures

Now we will focus on four of the factors we have discussed, which will help us to isolate five patterns of representation.

- Iconicity. Recall, we use this term to mean that the representational *elements* have richly grounded meaning. We distinguished fully iconic, partly iconic and non-iconic systems of representation.
- Systematic, constrained interpretation. In both the picture and the map, spatial relations among representations correspond to spatial relations among the represented items. Spatial relations

are not represented by individual symbols, as in text, but through a system of relationships among representations. The relations among representations obey constraints that correspond to those obeyed by the relations they represent.

- Localization. In the picture and the map all the representing facts that carry information about a given individual object will involve a single (token) representation of that object. It is fully localized. With charts there are some constraints, although more than one token is allowed; they are partially localized. There are no constraints with text; the information is not localized.
- Berkeley determinateness. This is systematic homomorphism when a determinate interpretation is provided, e.g. a scale and a north arrow for maps.

We can depict these factors in a way that gives us four rough but useful categories, going from text-like to picture-like.

	Iconic	Systematic and Constrained Interpretation	Berkeley Determinate	Localized
Texts	No	No	No	No
Charts	No	No	No	Partly
Diagrams	Partly	Yes	No	Fully
Maps	Partly	Yes	Yes	Fully
Pictures	Fully	Yes	No	Fully

TABLE 1 Five types of representation

30.4 ASL as a Heterogeneous System of Communication

On our analysis of ASL, it has three modes or "states", that reflect the way that space is being used to carry information. We call these states *Text*, *Organization by Region* or *OR* and *Stage*. We claim that the text state is text-like, the OR state is chart-like and the Stage state is map-like, in the senses indicated by the chart in Table 1.

Text. This is ordinary signing space. Gestures that correspond to vocabulary items are strung together to make longer phrases and sentences. There may be more iconicity involved than is typical in ordinary spoken languages, but meaning is basically a matter of arbitrary conventions. Representation is not systematic and constrained, but piecemeal and unconstrained in the way that speech or written text is.

ASL in text state is as expressive as any natural language. Abstract ideas, disjunctions, and the like can all be expressed; its limitations at any given time are just the limits of the vocabulary—as with any relatively small language, the development of technical terminology may lag from time to time.

One problem with ASL in text state, however, is that it can be very slow. Bellugi has shown that, although on average it takes twice as long to make a sign as to utter the corresponding English word, signers stay even with speakers in the rate at which they express thoughts (Klima and Bellugi 1979). An important reason for this is the use of the other two states.

The Stage. In the talk on which this paper is based, we demonstrated the use of the stage, but description will have to suffice here. Our favorite illustration is based on an actual event, a small automobile accident in which a member of our group was involved. Here is a textual description:

> I was stopped at light, thinking about nothing in particular. Suddely, a car ran into my car from the right rear. It scraped along the right side of my car, knocking it to one side, and came to a stop ahead and to the right of my car.

A signer could do this just like it is done in English, with signs corresponding one-to-one with words. That's not how it would normally be done in ASL—and in fact if one tried to do it that way, it would be thought of as something quite different, "Signed Exact English". The normal way to convey this description would involve using the hands to demonstrate the crash. The hands would be held in a shape that conventionally means "vehicle", but the movements are homomorphic and readily inferrable. The story can be told in about the same time it takes to tell it in spoken English, and in far less time than it would take to tell it in Signed Exact English.

The stage is a bit unlike any of our types of representation, because it is dynamic; it should be compared to a motion picture, not a still one. But we will set that important point aside for the time being. By our criteria, the stage is not picture-like because it is not fully iconic. It is localized. The hands (or particular fingers—depending) represent individual objects, and any individual object will be represented at any given time only by one hand or finger. It is systematic and constrained. Spatial relations between the representing hands or fingers represent spatial relations between the represented objects; movement means movment, and so forth. It need not be determinate. So on our criteria the stage is diagram-like.

The stage has the disadvantages of diagram-like representation as well as the advantages. Abstract ideas and disjunctions are not easily representable.

Organization by Regions. The third state uses localization to group information about a given individual or topic, but does not use space homomorphically.

Suppose you are comparing Neil and George. Neil is medium-height; George is tall. Niel is bubbly; George is reserved. And so forth. Rather than simply stay in text state, an ASL signer would subdivide the signing space into two regions, one for Neil and one for George. Then she would sign the various attributes, without repeating the subject term, indicating the subject by the region where the signing took place. Assuming Neil and George were both thirty-five, she could make the sign for that age in a neutral space, and then use the bi-directional sign "same" between the regions.

This state of ASL meets our criteria for being chart-like. It is non-iconic, not systematic and constrained and not determinate. But it differs from unconstrained text in using location to group information, and obeying the unique-token requirement.[6]

State-transitions. Since space is used differently depending on whether the signer is in the text, stage or OR state, there is the possibility of ambiguity and confusion. In the example above, it would not have been implied that George and Neil were or ever had been beside each other. But locating the two signs in the same relative positions to one another, while in the stage state, could have indicated this.

Actually, such confusion does not often arise. Signers and their audiences keep track of what state they are in, and what spatial relations signify. One reason for this is that there are definite rules for leaving one state and entering another. Suppose for example that the signer has introduced George and Neil into the discussion while in text state. She desires to do a comparison in a rapid and efficient manner, using the OR state. She would perform an operation somewhat analogous to "case-splitting" in Hyperproof. In normal signing space, she could simply write the names on top of each other—the spatial relation would have no meaning. In OR state, however, she shifts her body slightly to the right, and fingerspells "George" to the right of her normal fingerspelling space, a bit as if she were starting a list on a blackboard. Then she turns a bit to the left and similarly fingerspells "Neil" (or she could do this with the left hand). This clearly indicates a transition into the OR state.

[6]This is not exactly right, since Neil's name might not only be used to label his chart, but could conceivably also appear in the list of George's attributes—if, for example, Neil were George's brother

30.5 Conclusion

We have investigated factors that make representations more or less picture-like; several factors emerged, and consideration of those factors gave us five kinds of representational systems. We argued that ASL is a heterogenous system of communication, using three of the five types.

In the paper that forms the second part of this essay, we will develop an account of the devices used in ASL to move from one state of representation to another, and to transfer information across the states. We then look at the different kinds of representational systems that are used by modern, graphically-based computer interfaces. We examine how some of the same problems arise that have been successfully solved by ASL, and explore the possiblity of exporting such solutions from ASL to graphical interfaces. And we use our categories to speculate on the use of sound icons or earcons, three-dimensional sound, and other devices to create interfaces that provide the advantages of graphical interfaces to individuals who are blind.

References

Barwise, Jon, and John Etchemendy. 1990a. Information, Infons, and Inference. In *Situation Theory and Its Applications, I*, ed. Robin Cooper, Kuniaki Mukai, and John Perry. CSLI Lecture Notes Number 22. Stanford: CSLI Publications.

Barwise, Jon, and John Etchemendy 1990b. Visual Information and Valid Reasoning. In *Visualization in Mathematics*, ed. W. Zimmerman. Washington, DC: Mathematical Association of America.

Barwise, Jon, and John Etchemendy. 1994. *Hyperproof.* Stanford: CSLI Publications.

Etchemendy, John. 1976. *The Theory of Propositions in Wittgenstein's Tracatus.* Masters Thesis. Reno, Nevada: Department of Philosophy, Univeristy of Nevada.

Frishburg, N. 1975. Arbitrariness and iconicity: Historical change in American Sign Language. *Language* 51:676–710.

Klima, E., and U. Bellugi. 1979. *The Signs of Language.* Cambridge, Mass.: Harvard University Press.

Larkin, J.H., and H.A. Simon. 1987. Why A Diagram is (sometimes) Worth Ten Thousand Words. *Cognitive Science* 11:65–100.

Lewis, D. 1969. *Convention: A Philosphical Study.* Cambridge, Mass.: Harvard University Press.

Macken, Elizabeth, John Perry, and Cathy Haas. 1993. Richly Grounded Symbols in American Sign Language. *Sign Language Studies* 81:375–394.

Stenning, Keith, and Jon Oberlander. 1991. Words, Pictures and Calculi. In *Situation Theory and Its Applications, 2*, ed. Gordon Plotkin Jon Barwise, Jean Mark Gawron and Syun Tutiya. CSLI Lecture Notes Number 26. Stanford, Ca.: CSLI Publications.

Wittgenstein, Ludwig. 1922. *Tractatus-Logico Philosophicus.* . London: Routledge and Kegan Paul.

31

Information Flow and Relevant Logics

Greg Restall

Introduction

John Perry, one of the two founders of the field of situation semantics, indicated in an interview in 1986 that there is some kind of connection between relevant logic and situation semantics.

> I do know that a lot of ideas that seemed off the wall when I first encountered them years ago now seem pretty sensible. One example that our commentators don't mention is relevance logic; there are a lot of themes in that literature that bear on the themes we mention. (Barwise and Perry 1985)

In 1992, in *Entailment* volume 2, Nuel Belnap and J. Michael Dunn hinted at similar ideas. Referring to situation semantics, they wrote

> ... we do not mean to claim too much here. The Barwise-Perry semantics is clearly independent and its application to natural-language constructions is rich and novel. But we like to think that at least first degree (relevant) entailments have a home there. (Anderson et al. 1992)

In this paper I show that these hints and gestures are true. And perhaps truer than those that made them thought at the time.

Thanks to Jon Barwise for making his paper accessible over the Internet before its publication. Thanks too to Tom Burke, David Israel, Chris Menzel, John Perry, László Pólos, and an audience at the *Automated Reasoning Project* for helpful comments, encouragement and criticism.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

31.1 Routley-Meyer Frames

Relevant logics[1] are interesting things. One motivating principle for these logics is the requirement that if a conditional of the form $\phi \to \psi$ is to be true, there must be a connection between ϕ and ψ. The antecedent ϕ must somehow be *relevant* to the consequent ψ. This is simple to chart proof-theoretically, and for years, this was how relevant logics were studied. (A great deal of that work is charted in Anderson and Belnap 1975.)

It was quite a deal harder to give relevant logics a *semantics* like the possible-worlds semantics for modal logics. There is a good reason for this. In possible worlds talk, a conditional $\phi \to \psi$ is true at a world w (say $w \models \phi \to \psi$) if and only if for each world v accessible from w (say wRv) if ϕ is true at v then ψ is also true at v. In other words, a conditional is true at a world if and only if it is truth preserving at all accessible worlds. On this story tautologous conditional, like $\phi \to \phi$, is true at every world. This means that any conditional with a tautologous consequent, like $\psi \to (\phi \to \phi)$, is *also* true at every world. But this is a paradigmatic case of an irrelevant conditional (when ϕ has nothing to do with ψ). This happens *whatever* the accessibility relation R is like. So a possible-worlds semantics for relevant logics is out of the question.

Well, *not quite*. The answer to the problem is to liberate the accessibility relation. Routley-Meyer frames use a *ternary* relation to model the conditional. The clause for conditionals is this:

$$w \models \phi \to \psi \text{ if and only if for each } u, v \text{ where } Ruwv, \text{ if } u \models \phi \text{ then } v \models \psi.$$

And this does the trick.[2] We can have worlds where $w \not\models \phi \to \phi$, simply by having $Ruwv$ for some worlds u and v where $u \models \phi$ and $v \not\models \phi$. This is not hard to do, formally speaking. More bells and whistles are needed to get the semantics up and running. Details can be found in many places (Anderson et al. 1992, Priest and Sylvan 1992, Restall 1993, Restall 1994a, Routley and Meyer 1973 and Routley et al. 1982) but all presentations are *something* like this.

Definition 1 A *Routley-Meyer frame* is a 4-tuple $\langle g, W, R, \sqsubseteq\rangle$ satisfying the following conditions:

- $\sqsubseteq$ is a partial order on W.
- If $Rwuv$ and if $w' \sqsubseteq w, u' \sqsubseteq u$ and $v \sqsubseteq v'$ then $Rw'u'v'$ also.

[1] I am an Australian, and Australians tend to call these logics "relevan*t* logics," as opposed to the American "relevan*ce* logics." There is a long and involved story behind this geographical bifurcation of terminology. A story into which we will not go at the moment.

[2] In the literature, the first two places of the relation R are often swapped. (So, it would be '$Rwuv$' and not '$Ruwv$' in the clause displayed.) We use this arrangement for a smooth transition to what follows.

- $g \in W$. $Rwgv$ if and only if $w \sqsubseteq v$.

Definition 2 An *evaluation* $\models$ on a Routley-Meyer frame $\langle g, W, R, \sqsubseteq \rangle$ is a relation between worlds and formulae such that

- If $w \sqsubseteq v$ then if $w \models p$ then $v \models p$ for atomic propositions p.
- $w \models \phi \wedge \psi$ if and only if $w \models \phi$ and $w \models \psi$.
- $w \models \phi \vee \psi$ if and only if $w \models \phi$ or $w \models \psi$.
- $w \models \phi \to \psi$ if and only if for each u, v where $Ruwv$, if $u \models \phi$ then $v \models \psi$.

In the semantics a theorem is simply something true at g in all evaluations on all frames. (There's nothing true *everywhere* in all frames, so we need a special world g to record logical consequence, where logical consequence is still preservation across all worlds.) It is simple to show that these

$$\phi \to \phi \vee \psi \quad \phi \wedge \psi \to \phi \quad (\phi \to \psi) \wedge (\phi \to \theta) \to (\phi \to \psi \wedge \theta)$$

$$(\phi \to \theta) \wedge (\psi \to \theta) \to (\phi \vee \psi \to \theta) \quad \phi \wedge (\psi \vee \theta) \to (\phi \wedge \psi) \vee (\phi \wedge \theta)$$

are theorems, whereas these

$$\phi \to (\psi \to \psi) \qquad \phi \to (\psi \to \phi)$$

are not. In fact, many more things are not theorems as the semantics stands. There are counterexamples to each of these

$$\phi \to \big((\phi \to \psi) \to \psi\big) \quad \phi \wedge (\phi \to \psi) \to \psi$$

$$(\phi \to \psi) \to \big((\psi \to \theta) \to (\phi \to \theta)\big) \quad (\phi \to \psi) \to \big((\theta \to \phi) \to (\theta \to \psi)\big)$$

So, the logic is quite weak insofar as theorems relating to the conditional. This is as one would expect, for there are few conditions on the relation R. If we add more conditions on R, we get more theorems.

There are a number of unanalysed concepts in this presentation of the semantics for relevant logics. It is hard to understand what the worlds are, what a 'logic' world could be, what the relation R is grounded in, and so on. There are many different interpretations offered, but none have very widespread support. This is a thorny problem for practitioners of relevant logic, and it is one place where recent work in situation semantics can help.

31.2 Information Structures

In Barwise's recent work on information flow in situation semantics, he constructs a formalism which we can use to describe information flow (Barwise 1993). The crucial components are *situations* (with which you are familiar) and *channels*. Channels relate situations to one another, and they ground the flow of information. We classify situations with *types*, and we classify channels with *constraints*, which are of the form $\phi \to \psi$ where

ϕ and ψ are types. The details of the definition of an information structure are given below.

Definition 3 An *information structure* $\langle S, T, C, \models, \mapsto, \|, ; \rangle$ is a structure made up of a set S of *situations*, T of *types*, C of *channels*, a binary relation $\models$ which relates both pairs of situations and types and pairs of channels and constraints and a ternary relation $\mapsto$ relating channels to pairs of situations. (Here $s_1 \overset{c}{\mapsto} s_2$ is read as "c is a channel from s_1 to s_2.") This structure must satisfy a number of further conditions:

- Types are closed under binary operations $\wedge$ and $\vee$. Furthermore, for each $s \in S$ and each $\phi, \psi \in T$, $s \models \phi \wedge \psi$ if and only if $s \models \phi$ and $s \models \psi$, and $s \models \phi \vee \psi$ if and only if $s \models \phi$ or $s \models \psi$.
- For every $\phi, \psi \in T$, the object $\phi \to \psi$ is a *constraint*. The relation $\models$ is extended to channels and constraints in the way indicated: $c \models \phi \to \psi$ if and only if for each $s_1, s_2 \in S$ where $s_1 \overset{c}{\mapsto} s_2$, if $s_1 \models \phi$ then $s_2 \models \psi$. (So, a channel supports a constraint, just when for each pair of situations s_1, s_2 related by the channel, if s_1 supports the antecedent, s_2 supports the consequent. This is the crux of information flow.)
- There is a *logic channel*, channel 1. It relates each situation s to all situations that contain it. In other words, $s \overset{1}{\mapsto} s'$ if and only if $s \sqsubseteq s'$, where $s \sqsubseteq s'$ if and only if whenever $s \models \phi$, $s' \models \phi$ too.
- Every pair of channels c_1 and c_2 has a unique *sequential composition* $c_1; c_2$ (such that $s_1 \overset{c_1;c_2}{\longmapsto} s_2$ if and only if there is a situation s such that $s_1 \overset{c_1}{\mapsto} s$ and $s \overset{c_2}{\mapsto} s_2$). In addition, $c_1; (c_2; c_3) = (c_1; c_2); c_3$. Sequential composition enables us to link together long chains of information transfers.
- Every pair of channels c_1 and c_2 has a unique *parallel composition* $c_1 \| c_2$ (such that $s_1 \overset{c_1 \| c_2}{\longmapsto} s_2$ if and only if $s_1 \overset{c_1}{\mapsto} s_2$ and $s_1 \overset{c_2}{\mapsto} s_2$). In addition, parallel composition is commutative, associative, and idempotent. Parallel composition gives us a way to add together information transfers from the same signal (antecedent) to the same target (consequent).
- Channel c_1 is a *refinement* of channel c_2, written $c_1 \preceq c_2$ iff $c_1 = c_1 \| c_2$. So, a channel c_1 is a refinement of c_2 just when it relates *fewer* pairs. This means it will support more constrains by being more discriminating in what it relates.
- Sequential composition preserves refinement. That is, if $c_1 \preceq c_2$ then $c_1; d \preceq c_2; d$ and $d; c_1 \preceq d; c_2$.

Recall the treatment of declarative utterances on Barwise's approach. A declarative utterance has both a demonstrative and a declarative content —

respectively, a situation and a type. We pick out a situation and classify it with a type whenever we make a declarative utterance. According to Barwise's work on conditionals, whenever we utter a conditional, we classify a *channel* with a *constraint*. This makes sense, because when I utter a conditional, like

If white exchanges knights on *d*5 she will lose a pawn.

in the course of a game of chess, I am stating that there is some kind of relation between antecedent and consequent situations. There is a channel, relating next-move situations to situations which follow from it in the course of the game. I am saying that in all of the antecedent situations in which white exchanges knights on *d*5, in the consequent situations, she loses a pawn. This is exactly what the channel-constraint evaluation clause dictates.

In Barwise's initial account of information flow, he shows how a range of model structures in logic, computer science and information theory can each be seen as models of information flow. We'll show that a large class of Routley-Meyer frames also count as models of information flow.

31.3 Frames Model Information Flow

Recall the condition for a channel to support a conditional type.

The channel $c \models \phi \to \psi$ if and only if for all situations s_1, s_2, if $s_1 \overset{c}{\mapsto} s_2$ and $s_1 \models \phi$ then $s_2 \models \psi$.

Clearly this is reminiscent of the modelling condition of conditionals in frame semantics. If we take it that channels *are* situations, then the condition *is* that of the conditional in the frame semantics, where $\mapsto$ is R.

In frame semantics $x \overset{y}{\mapsto} z$ means that the conditional information given by y applied to x results in no more than z. This grounds the monotonicity condition

$$\text{If } x' \sqsubseteq x,\ y' \sqsubseteq y \text{ and } z \sqsubseteq z' \text{ and } x \overset{y}{\mapsto} z \text{ then } x' \overset{y'}{\mapsto} z'.$$

It is natural to take the serial composition $x; y$ to be contained in situation z just when $x \overset{y}{\mapsto} z$. This is because $x \overset{y}{\mapsto} z$ is read as "applying the information in y to that in x gives information in z." But serially composing x and y is just applying the information from y to that in x in order to get a new channel. So, if the application of y to x is bounded above by z, we must have $x; y$ contained in z (given the identification of channels and situations). And *vice versa*. So, from now we will read $x \overset{y}{\mapsto} z$ as $x; y \sqsubseteq z$ and *vice versa*.

What does associativity of channels mean in this context? We simply require that $(x; y); z \sqsubseteq u$ if and only if $x; (y; z) \sqsubseteq u$ for each u. But this comes out as follows. $(x; y); z \sqsubseteq u$ if and only if for some v, $x; y \sqsubseteq v$ and $v; z \sqsubseteq u$. In other words, for some v, $x \overset{y}{\mapsto} v$ and $v \overset{z}{\mapsto} u$. Conversely,

$x;(y;z) \sqsubseteq u$ if and only if for some w, $x;w \sqsubseteq u$ and $y;z \sqsubseteq w$, which can be rephrased as $x \overset{w}{\mapsto} u$ and $y \overset{z}{\mapsto} w$. Given our rewriting of sequential channel composition in terms of the channel relation $\mapsto$ we have an associativity condition in terms of $\mapsto$ alone. This will be enough to start our definition of a frame modelling information flow.

Definition 4 A *bare frame* is a quadruple $\langle g, S, \mapsto, \sqsubseteq\rangle$, where S is a set of *situations*, $\mapsto$ is a ternary relation on S, $g \in S$, is the *logic situation* and $\sqsubseteq$ is a partial order on S. The objects satisfy the following further conditions.

- If $x' \sqsubseteq x$, $y' \sqsubseteq y$ and $z \sqsubseteq z'$ and $x \overset{y}{\mapsto} z$ then $x' \overset{y'}{\mapsto} z'$
- $x \overset{g}{\mapsto} y$ if and only if $x \sqsubseteq y$.
- $(\exists v)(x \overset{y}{\mapsto} v$ and $v \overset{z}{\mapsto} u)$ if and only if $(\exists w)(x \overset{w}{\mapsto} u$ and $y \overset{z}{\mapsto} w)$.

Now that we have the structures defined, we need to show that these structures really model the axioms, by defining parallel and serial composition.

Take situations a and b. Their serial composition ought to be the 'smallest' situation x (under $\sqsubseteq$) such that $a \overset{b}{\mapsto} x$ given our motivation of identifying $a \overset{b}{\mapsto} c$ with $a;b \sqsubseteq c$. However, nothing assures us that such a minimal situation exists. There may be two candidate situations which agree with regard to all conditionals, but disagree with regard to a disjunction $p \vee q$. As situations are prime, neither of these is minimal. Instead of requiring that such a situation exist, we will model the serial composition of these two situations as the *set* $\{x : a \overset{b}{\mapsto} x\}$. If we take a set to support the type ϕ just when *each* of its elements supports ϕ, the set $\{x : a \overset{b}{\mapsto} x\}$ will work as the serial composition of a and b. It may be considered to be a 'non-prime situation,' or merely as the information shared by a collection of situations. From now on we take our channels to be sets of situations like this. A channel can be taken to be part of a situation just when the situation is an element of the channel. Let's make things formal with a few definitions.

Definition 5 Given a bare frame $\langle g, S, \mapsto, \sqsubseteq\rangle$

- $X \subseteq S$ is a *cone* iff for each $x \in X$, if $x \sqsubseteq y$ then $y \in X$.
- If X is a cone, $X \models \phi$ iff $x \models \phi$ for each $x \in X$.
- If X, Y and Z are cones, $X \overset{Y}{\mapsto} Z$ if and only if for every $z \in Z$ there are $x \in X$ and $y \in Y$ where $x \overset{y}{\mapsto} z$.
- If X and Y are cones, $X \sqsubseteq Y$ if and only if $Y \subseteq X$. In addition, $X;Y = \{z : X \overset{Y}{\mapsto} z\}$, $X \parallel Y = \{z : X \sqsubseteq z$ and $Y \sqsubseteq z\}$.
- For each situation x, $\uparrow x$ is the principal cone on x. In other words, $\uparrow x = \{x' : x \sqsubseteq x'\}$.

Given these definitions, it is not difficult to prove the following results.

Lemma 1 *Given a bare frame $\langle g, S, \mapsto, \sqsubseteq\rangle$ with an evaluation $\models$*

- *$X \models \phi \to \psi$ iff for each pair of cones Y, Z, where $Y \overset{X}{\mapsto} Z$, if $Y \models \phi$ then $Z \models \psi$.*
- *$X \models \phi \to \psi$ iff for each pair of situations y, z, where $\uparrow y \overset{X}{\mapsto} \uparrow z$, if $y \models \phi$ then $z \models \psi$.*
- *$X \models \phi \wedge \psi$ iff $X \models \phi$ and $X \models \psi$.*
- *$X \models \phi \vee \psi$ iff for each $x \in X$, either $x \models \phi$ or $x \models \psi$.*
- *$\uparrow x \sqsubseteq Y$ iff for each $y \in Y$, $x \sqsubseteq y$.*
- *$X \sqsubseteq \uparrow y$ iff $y \in X$*
- *$(\exists v)(X \overset{Y}{\mapsto} v$ and $v \overset{Z}{\mapsto} U)$ iff $(\exists w)(X \overset{w}{\mapsto} U$ and $Y \overset{Z}{\mapsto} w)$.*
- *$\uparrow x \sqsubseteq \uparrow y$ iff $x \sqsubseteq y$*
- *$\uparrow x \overset{\uparrow y}{\mapsto} \uparrow z$ iff $x \overset{y}{\mapsto} z$.*
- *$\uparrow x \models \phi$ if and only if $x \models \psi$.*

Proof. Straight from the definitions. We leave them as an exercise. □

Because of the last three results in that lemma, principal cones will do for situations whenever they occur. From now, we will slip between a principal cone and its situation without mentioning it.

The significant result is that $X;Y$ really is the serial composition of X and Y. In other words, we can prove the following:

Lemma 2 *For all cones X and Y, and for all situations a and c, $a \overset{X,Y}{\longmapsto} c$ iff there is a situation b such that $a \overset{X}{\mapsto} b$ and $b \overset{Y}{\mapsto} c$.*

Proof. Suppose that $a \overset{X,Y}{\longmapsto} c$. Then for some $d \in X;Y$, $a \overset{d}{\mapsto} c$. However, if $d \in X;Y$ we must have an $x \in X$ and a $y \in Y$ where $x \overset{y}{\mapsto} d$. So, $x \overset{y}{\mapsto} d$ and $a \overset{d}{\mapsto} c$. This means that for some b, $a \overset{x}{\mapsto} b$ and $b \overset{y}{\mapsto} c$ by one half of the associativity condition This means that $a \overset{X}{\mapsto} b$ and $b \overset{Y}{\mapsto} c$ as desired.

The converse merely runs this proof backwards and we leave it as an exercise. □

Then serial composition is associative because of our transitivity condition on modelling conditions.

Lemma 3 *In any bare frame, for any cones X, Y and Z, $X;(Y;Z) = (X;Y);Z$*

Proof. If $w \in X;(Y;Z)$ then $X \overset{Y,Z}{\longmapsto} w$, which means that for some $x \in X$ and $v \in Y;Z$, $x \overset{v}{\mapsto} w$. Similarly, $v \in Y;Z$ means that for some $y \in Y$ and $z \in Z$, $y \overset{z}{\mapsto} v$. But this means that for some u, $x \overset{y}{\mapsto} u$ and $u \overset{z}{\mapsto} w$, which gives $u \in X;Y$ and so, $w \in (X;Y);Z$ as desired. The converse proof is the proof run backwards, as usual. □

As things stand, parallel composition may not work as we intend. We may have cones X and Y for which the parallel composition is empty (that is, there is no z where $X \sqsubseteq z$ and $Y \sqsubseteq z$), but this means that $a \overset{X\|Y}{\longmapsto} b$ does not hold (look at the clauses). However, we may still have $a \overset{X}{\mapsto} b$ and $a \overset{Y}{\mapsto} b$. We need an extra modelling condition.

Definition 6 A bare frame $\langle g, S, \mapsto, \sqsubseteq \rangle$ is an *information frame* if and only if for each a, b, x, y where $a \overset{x}{\mapsto} b$ and $a \overset{y}{\mapsto} b$, there is a z where $x \sqsubseteq z$ and $y \sqsubseteq z$ and $a \overset{z}{\mapsto} b$. Parallel composition in an information frame is defined as you would expect. $X \parallel Y = \{z : X \sqsubseteq z \text{ and } Y \sqsubseteq z\}$.

Lemma 4 *In any information frame, for all cones X and Y, and for all situations a and b, $a \overset{X}{\mapsto} b$ and $a \overset{Y}{\mapsto} b$ iff $a \overset{X \parallel Y}{\longmapsto} b$. Furthermore, parallel composition then does what we would expect of it: $X \parallel X = X$, $X \parallel Y = Y \parallel X$ and $X \parallel (Y \parallel Z) = (X \parallel Y) \parallel Z$. In fact $X \parallel Y = X \cap Y$.*

Proof. Trivial. Clearly, if $a \overset{X \parallel Y}{\longmapsto} b$ then $a \overset{X}{\mapsto} b$ and $a \overset{Y}{\mapsto} b$ by monotonicity. The converse holds by the definition of an information frame. The properties of information frames follow from the fact that $\sqsubseteq$ is a partial order. The fact that parallel composition turns out to be intersection follows from the fact that it is defined on cones on the partial order. □

The relation $\preceq$ of refinement is simply $\sqsupseteq$; $X \preceq Y$ iff $X \parallel Y = X$, iff $X \cap Y = X$ iff $X \subseteq Y$ iff $X \sqsupseteq Y$.

The last thing to show is that serial composition preserves refinement. We need to show the following:

Lemma 5 *In any information frame, if $X_1 \sqsubseteq X_2$ and $Y_1 \sqsubseteq Y_2$ then $X_1; Y_1 \sqsubseteq X_2; Y_2$.*

Proof. This follows immediately from the hereditary conditions on $\mapsto$. □

So, our frame structures, with a few extra bells and whistles, are models of information flow. The resulting logic is weaker than the relevant logic **R** (and it even invalidates some theorems of **E**), but it is nonetheless quite an interesting system in its own right. Elsewhere (Restall 1994b) I show it (and several of its neighbours) to be decidable, and naturally motivated from other points of view.

31.4 Frames Are the Best Model of Information Flow

Barwise's paper shows that model structures like those of classical logic and intuitionistic logic also model information flow. If we were only able to put relevant logics into this class, this would not say very much about them. Why favour systems like relevant logics? Why favour the distinctions that frame semantics can draw (between serial and parallel composition) and the identifications it makes (such as identifying channels with situations)?

31.4.1 Why Frames?

Firstly, frames are the sensible model of information flow, because they remain faithful to the original intuitions about the meanings of conditional utterances. If we say that situations and channels are totally distinct, then we have a problem. We decided that a conditional statement "if S_1 then S_2" has the constraint $\phi \rightarrow \psi$ as descriptive content, and a particular *channel* c as demonstrative content. But we already decided that other statements have *situations* as their demonstrative content. Why are conditional statements any different? To avoid such an arbitrary distinction, we must admit a relationship between situations and channels. If the conditional "if S_1 then S_2" has a situation s as its demonstrative content, then we may take s itself to be a channel between situations. This is admitted in the frame semantics. Each situation is a channel and arbitrary *cones* of situations are also channels — chiefly to deal with channel composition.

In other words, once we recall a major motivating application of information flow (modelling conditionals in situation semantics in terms of regularities grounded in the world) the frame semantics identification of channels with situations is the natural conclusion.

31.4.2 Why So Many Frames?

We know that frames model information flow. It doesn't follow that every information frame is relevant to our cause. Barwise specified that serial composition be associative, leaving open whether it was symmetric, idempotent or whatever. Our formalism leaves it open for us to construct frames in which serial composition fails to be symmetric or idempotent. Is this justified, or should there be more conditions on frames?

To deal with a part of this question, let's see how a failure of the idempotence of composition can be motivated.[3] We want to find a counterexample to $x; x \sqsubseteq x$. (It is simple to show that in all of Barwise's examples in his paper, this condition holds. Yet, it does not hold in all information structures.) But this is simple. For many channels, you may eke out more information by repeated application. We need only find a domain in which using a channel twice (serially) yields more information than using it once. We'll just sketch one such application, and leave the reader to think of more.

Take situations to be mathematical proofs, and we will take the information content to be the *explicit content* of the proof.[4] That is, the things that are stated in the proof. We can model the information flow from a proof to another proof (which may contain the first proof as a part) by way

[3] I am especially fond of this example, for a great deal of my research has been into logics in which this rule fails (Restall 1994a). It is equivalent to rejecting $\phi \wedge (\phi \rightarrow \psi) \rightarrow \psi$ and $\big(\phi \rightarrow (\phi \rightarrow \psi)\big) \rightarrow \big(\phi \rightarrow \psi\big)$.

[4] And its real logical consequences, as defined in the situation-theoretic account. So, if $A \wedge B$ is a part of the information content of a proof, so are A, B and $A \vee C$.

of information links that relate proof-situations by means of the deductive principles of mathematics. For example, one such rule is this:

If n is even, so is $n + 2$.

If this rule is warranted by a channel x, then if y and z are proof situations where $y \overset{x}{\mapsto} z$, and y supports the claim '6 is even,' then z supports the claim '8 is even.' The proof z may have been produced from the proof y by applying the rule of inference we've seen, and the channel x indicates (or warrants) that application.

Now a proof may have '8 is even' as a part of its explicit content without '10 is even' also being a part of its content. To get that from the initial proof situation y, we need to apply the rule *twice*, and so, use the channel x twice. The channel $x;x$ therefore warrants more than x. We may have $y \overset{x;x}{\longmapsto} z$ without $y \overset{x}{\mapsto} z$.

31.5 Putting the Account to Work

Now that we have the formalism in place, we can describe the relationship between channels (or situations) and conditionals. This will also give the wherewithal to give an account of some seemingly paradoxical arguments using conditionals. This is important, because while relevant logics are good at blocking paradoxes such as $\phi \to (\psi \to \psi)$ and so on, they are not so good at blocking other difficulties with conditionals. Common examples are failures of transitivity (from $\phi \to \psi$ and $\psi \to \theta$ to deduce $\phi \to \theta$) and strengthening of the antecedent (from $\phi \to \psi$ to deduce $\phi \wedge \theta \to \psi$). The formal system we consider has

$$(\psi \to \theta) \to \big((\phi \to \psi) \to (\phi \to \theta)\big) \qquad (\phi \to \psi) \to (\phi \wedge \theta \to \psi)$$

as theorems. How can we avoid being nailed by the counterexamples people give?

The situated semantics of natural language provides an answer: We must recall that all declarative utterances have a demonstrative content (a situation) and a descriptive content (a type). This is no different for conditional utterances, which have a situation as demonstrative content, and a type of the form $\phi \to \psi$ as descriptive content. The truth or otherwise of the conditional depends on whether the situation described is of the type or not. This situation relativity gives us the means to give an account of odd-sounding arguments.

The first is a putative counterexample to transitivity. Consider the two conditionals:[5]

If an election is held on December the 25th, it will be held in December.

[5] In Australia the timing of elections is decided by the government of the day.

If an election is held in December, it will not be held on December the 25th.

By a naïve application of transitivity we could argue from these two claims that if an election is held on December the 25th, it will not be held on December the 25th. This is an odd conclusion to draw. The analysis in terms of channels and constraints can help explain the puzzle without rejecting transitivity.

Firstly, consider the situation described by the first conditional. This situation — in its job as a channel — pairs election-on-December-the-25th situations with election-in-December situations. Presumably this channel arises by means of logical truth, or our conventions regarding dates and months. Quite probably, it pairs each election situation with itself. If the antecedent situation has the election on December the 25th, then the consequent situation (the same one) has the election in December. There is little odd with this scenario.

The second situation pairs antecedent election-in-December situations with consequent election-not-on-December-the-25th situations. Given the plausible assumption that it pairs antecedent situations only with identical consequent situations (or at least, consequent situations not incompatible with the antecedent situations — so it will not pair an antecedent situation with a consequent situation in which the election occurs at a *different* date) it will 'filter out' antecedent situations in which the election is held on Christmas Day. In Barwise's words, these situations are *pseudo-signals* for the constraint. These aberrant situations are not related (by our channel) with any other situation at all. The channel only relates 'reasonably likely' election situations with themselves, and so, it supports the constraint that elections in December are not held on Christmas Day just because it doesn't relate those (unlikely) situations in which an election is held on that day.

Given these two channels, it is clear that their composition supports the constraint 'if there is an election on December the 25th, then the election is not on December the 25th' simply because there are no signal/target pairs for that channel in which the signal is a situation in which the election is on December the 25th. The composition of the channels filters out these antecedent situations by construction. That channel supports other odd constraints like 'if there is an election on December the 25th, then Queensland has won the Sheffield Shield for the last twenty years.' This does not tell us anything enlightening about what would happen were an election to actually occur on December the 25th — it only tells us that this particular channel has ruled that possibility out.

Composition of channels may 'filter out' possibilities (like elections held on Christmas day) that will later become relevant. Then we typically broaden the channels to admit more possibilities. (This is akin to expanding the set of 'nearby possible worlds' used to evaluate counterfactuals on the Lewis-Stalnaker accounts.) Typically when we utter a counterfactual conditional we mean that situations classified by the antecedent will feature in signal/target pairs of the channel being utilised. (Otherwise, why utter the conditional?) In cases like this argument, the composed channel is not like this. The antecedent situations being described do not feature as in signal-target pairs of the channel being classified. So, the conditional given by the Xerox principle is not the same as the conditional you would typically be expressing had you said

If an election is held on December the 25th, it will not be held on December the 25th.

Had you said that (and it is a strange thing to say) then most likely, the channel being classified would have as signal/target pairs some situations in which the election is held on December the 25th. (Otherwise, why call attention to their possibility?) And if this is so, the conditional you express by asserting the sentence will differ from that arising from the Xerox principle but this only points out the channel relativity of conditionals. This is parallel to the situation-theoretic fact that propositions are situation relative. So, the principle itself is sound, but difficulties like these must keep us alert to the way it is used.

As was the case with the transitivity, we can use the channel-theoretic account to explain the oddity of certain 'deductions' using conjunction. For example, given a number of background conditions

If it snows, I won't be surprised.
If it hails, I won't be surprised.

could be both true. It may be quite cold, and both snow and hail could be possible given the current weather conditions. Furthermore, my expectations are attuned to the relevant facts about the weather, and so, I won't be surprised under either of those conditions. However, in this situation, there is no guarantee that

If it both snows and hails, I won't be surprised.

because the combination of snow and hail is a very rare thing indeed — and I may be aware of this. This appears to be a counterexample to the principle of addition of information (collapsing the two conjuncts in the consequent to one). Yet, as with transitivity, this is not a real counterexample. What's more, the account in terms of channels can help us explain the surprising nature of the 'deduction.'

Consider the channels supporting the original two claims. Obviously they do not relate all snowing or hailing signal-target pairs with consequent mental states, because for some snowing or hailing situations (ones that are combined) I *am* surprised. So, the channels supporting these claims must ignore the possible (but rare) situations in which snow and hail coincides. In other words, snow-and-hail situations are pseudo-signals for this constraint. This is understandable, because it is a rare situation to encounter. Now when we consider the parallel composition of the two channels, it is simple to see that it has no signal-target pairs where it is snowing and hailing in the signal situation. In each of the original channels, these possibilities were filtered out, so they cannot re-emerge in their parallel composition. The third conditional is supported by the parallel composition of channels only vacuously. The composed channel does not relate any snowy-and-haily situations.

Were we to say 'if it both snows and hails, I won't be surprised' the channel classified would (usually) not be one that filters out odd snowy-and-haily situations, because we have explicitly mentioned that situation as a possibility. Again, we must be careful to not identify the conclusion of the addition of information principle with a claim we may express ourselves. For each declarative utterance, there is a corresponding situation or channel that is classified. Different utterances could well classify different situations or channels.

In this way we can use the formalism to explain the oddity in certain conditional argument forms. They are sensitive to the situations being described, which can vary from premise to conclusion, without this fact being explicit.

31.6 Metaphysical Issues

The semantic structure makes essential use of non-actual situations. Consider false conditionals with false antecedents. To make

If Queensland win the Sheffield Shield, grass will be blue

untrue at a situation x we need a y and z where $y \overset{x}{\mapsto} z$, $y \models$ Queensland win the Sheffield Shield and $z \not\models$ grass is blue. This requires the existence of a situation y that supports the claim that Queensland win the Sheffield Shield. This is patently false, so no actual situation supports it.[6] It follows that the conditional is true (which is an unsavoury conclusion) or there are some non-actual situations.

It follows that a decent account of conditionals in this formalisation must involve non-actual situations — so these don't qualify as bits of the world (in the same way that actual situations do, anyway). Barwise agrees. His answer to the question is that situations ought to be seen as

[6]This was false at the time of writing, but Queensland won its first ever Sheffield Shield in the summer of 1994–1995, rendering this example inappropriate. The reader can substitute another falsehood here.

mathematical objects for modelling possibilities, and not as real but not actual situations. (Barwise 1993)

But this won't suffice as an explanation — at least if this semantics is to work as an account of anything. Because if a *model* is to have any explanatory force, the things in the model must correspond to *something* real. If we have a mathematical model for a possibility, then if it does any explanatory work (which it does) there must be something *real* that corresponds to it, and that grounds the explanation in the Way Things Are. If these things are not "real but not actual situations" it would be interesting to hear what they are. Giving an account of these that doesn't amount to realism about (particular) non-actual situations is exactly parallel to the task of giving a non-realist account of possible worlds. Calling them "mathematical models" is honest, but it only pushes the question back to those who want to know what the model actually *models*. For this approach to have any chance of working without committing us to modal realism, you must explicate the notion of "modelling possibilities." At face value this does seem to involve a relationship between models and the possibilities they purport to model. However, there may be another way to cash out the conception: we can say that x models a possibility (or *represents* a possibility) if x *would* model (or represent) a situation, were things different in some relevant respect. (Chris Menzel (Menzel 1990) has worked out this approach in the context of modal logic.) To make this work in our context we need to spell out what way things are allowed to vary, and be more specific about the representation relation. However, it is plausible that some explanation like this might work.

Taking this line would result in the analysis being circular in one sense. Cashing out the notion of representation requires using conditionality or possibility — so we will not be giving a reductive account of conditionals. On this semantics, possibility or conditionality will be primitives. However, they will be primitives that are closely associated with other concepts such as the channeling relation between situations, and this, as in all formal semantics, will give us a helpful regimentation of our intuitions about conditionals, and it will give us a new way to analyse their semantic content.

References

Anderson, Alan Ross, and Nuel D. Belnap. 1975. *Entailment: The Logic of Relevance and Necessity, Volume 1.* Princeton: Princeton University Press.

Anderson, Alan Ross, Nuel D. Belnap, and J. Michael Dunn. 1992. *Entailment: The Logic of Relevance and Necessity, Volume 2.* Princeton: Princeton University Press.

Barwise, Jon. 1993. Constraints, Channels and the Flow of Information. In *Situation Theory and Its Applications, 3*, ed. Aczel, Israel, Katagiri, and Peters. CSLI Lecture Notes Number 37. Stanford: CSLI Publications.

Barwise, Jon, and John Perry. 1985. Shifting Situations and Shaken Attitudes. *Linguistics and Philosophy* 8:105–161.

Menzel, Christopher. 1990. Actualism, Ontological Commitment, and Possible World Semantics. *Synthese* 85:255–289.

Priest, Graham, and Richard Sylvan. 1992. Simplified Semantics for Basic Relevant Logics. *Journal of Philosophical Logic* 21:217–232.

Restall, Greg. 1993. Simplified Semantics for Relevant Logics (and some of their rivals. *Journal of Philosophical Logic* 22:481–511.

Restall, Greg. 1994a. *On Logics Without Contraction.* Doctoral dissertation, The University of Queensland, January.

Restall, Greg. 1994b. A Useful Substructural Logic. *Bulletin of the Interest Group in Pure and Applied Logic* 2. To appear.

Routley, Richard, and Robert K. Meyer. 1973. Semantics of Entailment 1. In *Truth Syntax and Modality*, ed. Hugues Leblanc. 199–243. Amsterdam: North-Holland. Proceedings of the Temple University Conference on Alternative Semantics.

Routley, Richard, Val Plumwood, Robert K. Meyer, and Ross T. Brady. 1982. *Relevant Logics and their Rivals.* Ridgeview.

32

Attunement to Constraints in Nonmonotonic Reasoning

WILLIAM C. ROUNDS AND GUO-QIANG ZHANG

Introduction

This paper began as an attempt to use domain theory to throw some light on some of the semantic problems in the area of non-monotonic logic within artificial intelligence. After having thought about some of the issues involved, though, it seemed to us that the examples and methods we use might be of interest in a broader setting. We are offering the results of our investigations in that spirit, hoping that this will be a start on the problem of putting the work in standard AI approaches to non-monotonicity together with current work on information flow (Seligman and Barwise 1993.)

We are interested in the subject of default or "normal" inferencing. On the surface of things, this can be exemplified within propositional logic using a *non-monotonic consequence relation* $\alpha \mathrel{|\!\sim} \beta$ between sentences α and β. The typical gloss of this is "birds normally fly." So, in terms of (one version of) situation theory, one would take α and β as types, as in the "smoke means fire" paradigm of Devlin (Devlin 1991.) In later work, Barwise (Barwise 1989) suggests that this relation, as a relation between types, would be an *indicating relation*, to be accompanied by a corresponding *signaling relation* $s_1 \to s_2$, where s_1 is a "site" supporting α, and s_2 is a site supporting β. In later

Imagine for a moment that there are no exceptions to any rules. We can then use a "strict" indicating relation $\alpha \Rightarrow \beta$ to mean that any object of

Research sponsored by NSF Grant IRI-9120851. A very preliminary draft of the paper appeared in the collection of position papers for the First Workshop on Principles and Practices of Constraint Programming, Newport, RI, May 1993, sponsored by ONR.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

type α is also of type β. (The pair (α, β) is called a constraint.) In this case, we can take the signaling relation to be the identity on the objects (sites). Then we can have a flow of information about such objects: we can associate the new information that an object s is of type β given that it was of type α. So far this is pretty unexceptional, but the interesting point is that *even when there are exceptions*, we still use the identity signaling relation to gain information about objects. That is, if we know that birds normally fly, and only that Tweety is a bird, then we infer (although possibly incorrectly) that Tweety can fly. The point of default reasoning is that we use such information virtually all the time.

This scenario has of course been extensively studied in artificial intelligence. Barwise has also given a situation-theoretic treatment of it in (Barwise 1989, chapter 5), where the additional idea of *background type* was introduced. The idea there was that a background type N might stand for "the type of birds which are not like penguins, emus, and ostriches", so that conditional on the background type N, the type "bird" would correctly entail the type "fly". This idea was subsequently refined, both by Barwise and by Seligman (Seligman and J.Barwise 1993), into the idea of an information channel. In channel theory, pairs $s1 \to s2$ are *classified* by some channel as being of type $\alpha \Rightarrow \beta$ or not. If this classification holds, and $s1$ is in fact of type α, then we can infer that $s2$ is of type β. Exceptions can occur here: when, for example, the pair $s1 \to s2$ fails to be classified either positively or negatively. Cavedon (Cavedon 1994) has given a semantics of conditional logic and a semantics for defaults using these ideas directly.

In this paper, we would like to present another information-theoretic proposal to model the above scenario. It also involves the notion of background, but in a somewhat different way. We think of a background constraint as a *strict* constraint relative to which we add non-strict *default constraints* of the kind mentioned above. The kind of background constraint we have in mind is the strict constraint that penguins are birds, and do not fly.

We work in first-order logic, and we model background constraints as conditional first-order sentences. So "penguins are birds" is just given as the usual universally quantified sentence (Horn sentence) expressing the constraint. We use background constraints to govern partial models of first-order logic. These models are constructed using systems of default rules, as in the *default logic* of Reiter (Reiter 1980), but where Reiter's rules build logical theories, our rules build models. Our approach takes advantage of domain-theoretic notions. A system of default rules is a straightforward generalization of Scott's notion of information system (Scott 1982.) We think of these default systems as "programs" which are created by a reasoning agent to satisfy default constraints.

We show how to treat the standard penguin example in our system, and we give what we hope are some other amusing case studies. The interesting observation here, we feel, is that we can show a specific example of what Devlin calls "attunement" to background constraints. The idea is that a reasoner will create default systems in response to experience, in an effort to make default constraints into hard ones. But these systems of defaults can be giving a lot of non-information and even false information in a probabilistic sense. They, and the default constraints themselves, should be undergoing revision. We illustrate this by considering an anomaly of Poole (Poole 1989), related to the so-called "lottery paradox" (Jr. 1961); and we consider a more complex case involving the well-known (folk?) "Nixon Diamond."

To this end, we introduce a notion both of non-monotonic consequence $\mid\!\sim$ which can be used to state default constraints, and, in the finite case, a probabilistic notion of "degree of belief" useful in analyzing the examples.

32.1 Relations to Standard AI Approaches

Etherington (Etherington 1988) gave the first model-theoretic semantics to Reiter's default logic. This was a system based on *sets* of first-order models. Marek, Nerode, and Remmel (Marek et al. 1990), gave a semantics for nonmonotonic rule systems. They translated Reiter's default rules into finitary formulas of a certain special infinitary logic. Extensions – the central construct of Reiter's logic – are viewed as models for certain formulas encoding the existence of default derivations.

Our approach has certain commonalities with the Nerode, Marek, and Remmel theory, in that we view extensions model-theoretically. However, we use extensions as models for ordinary first-order logic, not the special logic used by Marek, Nerode and Remmel. It will also be clear that first-order logic is not the only possible logic for which default models could serve as a semantic space. But we concentrate on the first-order case, since that involves the use of constraints.

Our treatment also has the advantage that one can analyze default reasoning situations by working directly with models, as one does all the time in ordinary mathematical reasoning. This contrasts with the approaches of MNR and Etherington, where in the first case, the logic describes a proof theory, and in the second, where one works with sets of first order models as models for default logic. We hold the thesis that Reiter's default systems should be regarded, not as proof rules, but as *algorithms for the direct construction of partial models for some appropriate logic.* This is a simple and radical reconstruction of default reasoning. To give it a proper explication, though, we use domain-theoretic tools – information systems and

Scott domains, in particular, since in our view default reasoning is about what happens when information necessary to resolve a question is lacking.

As we have stated, we want our models to be governed by *constraints*, which in our setting are thought of as laws which govern the behavior of partial models, but which are in the background. We encode a constraint theory into the monotonic forcing relation $\vdash$ of a Scott information system appropriate for a first-order logic semantics. How to accomplish this encoding was not absolutely clear. One possibility is to use a generalization of information systems themselves, due to Saraswat and Panangaden (Saraswat et al. 1991), to the first-order case. We have determined, however, that such a move is unneccessary. We represent constraint theories as a special case of ordinary monotonic information systems.

The next problem is how to add a non-monotonic component to information systems. This we have done by simply adding default forcing rules to Scott's systems.

A final problem is how to use the domains we generate as models for first-order logic itself, and specifically, how to interpret negation. We have chosen a restricted, positive version of first-order logic, which only allows negation on atomic formulas. Then we introduce a notion $\mid\!\sim$ of *non-monotonic consequence* between sentences of first-order logic, as in Kraus, Lehmann, and Magidor (Kraus et al. 1990.) We say that in a default model structure, one sentence non-monotonically entails a second if the second holds in all extensions of "small" partial models of the first. Here "smallness" is interpreted with respect to the natural partial order associated with a Scott domain.

We then turn to the construction of degrees of belief using finite models. Finite default models are of course a special case of our theory. We can generalize the usual finite model theory to partial models, and can use default rules to assign probabilities to statements in FOL, representing an agent's degree of belief in certain situations obtaining. This gives a way of thinking about the usual "birds normally fly" as a pseudo-probabilistic statement. We illustrate this method in the resolution of an anomaly with standard default reasoning, due to Poole (Poole 1989.)

The paper is organized as follows. In section 2 we cover the basics of domain theory and information systems, introduce our non-monotonic generalization, and state a representation theorem for default domains. In Section 3 we show how to interpret first-order positive logic using default models. This is where constraints play a crucial role. Then in Section 4 we introduce our notion of conditional degree of belief, and treat our examples.

32.2 Default Domain Theory

32.2.1 Information Systems

We review Scott's representation of domains using *information systems*, which can be thought of as general concrete monotonic "rule systems" for building Scott domains.

Definition 1 An information system is a structure $\underline{A} = (A, \ Con, \ \vdash)$ where

- A is a countable set of tokens,
- $Con \subseteq \mathrm{Fin}(A)$, the consistent sets,
- $\vdash \subseteq Con \times A$, the entailment relation,

which satisfy

1. $X \subseteq Y \ \& \ Y \in Con \Rightarrow X \in Con$,
2. $a \in A \Rightarrow \{a\} \in Con$,
3. $X \vdash a \ \& \ X \in Con \Rightarrow X \cup \{a\} \in Con$,
4. $a \in X \ \& \ X \in Con \Rightarrow X \vdash a$,
5. $(\forall b \in Y. X \vdash b \ \& \ Y \vdash c) \Rightarrow X \vdash c$.

Example: Approximate real numbers. For tokens, take the set A of pairs of rationals $\langle q, r \rangle$, with $q \leq r$.

The idea is that a pair of rationals stands for the "proposition" that a yet to be determined *real number* is in the interval $[q, r]$ whose endpoints are given by the pair.

Define a finite set X of "intervals" to be in in *Con* if X is empty, or if the intersection of the "intervals" in X is nonempty. Then say that a set $X \vdash \langle q, r \rangle$ iff the intersection of all "intervals" in X is contained in the interval $[q, r]$. Note that there is only atomic structure to these propositions. We cannot negate them or disjoin them.

The representation of Scott domains uses the auxiliary construct of ideal elements.

Definition 2 An (ideal) element of an information system $\underline{A}$ is a set x of tokens which is

1. consistent: $X \subseteq x \Rightarrow X \in Con$,
2. closed under entailment: $X \subseteq x \ \& \ X \vdash a \Rightarrow a \in x$.

The collection of ideal elements of $\underline{A}$ is written $|\underline{A}|$.

Example. The ideal elements in our approximate real system are in 1-1 correspondence with the collection of closed real intervals $[x, y]$ with $x \leq y$. Although the collection of ideal elements is partially ordered by inclusion, the domain being described – intervals of reals – is partially ordered by reverse interval inclusion. The total or maximal elements in the domain correspond to "perfect" reals $[x, x]$. The bottom element is a special interval $(-\infty, \infty)$.

It can be easily checked that for any information system, the collection of ideal elements ordered by set inclusion forms a Scott domain. Conversely, every Scott domain is isomorphic to a domain of such ideal elements. These results are basic in domain theory, and have been generalized to other classes of complete partial orders by Zhang (Zhang 1991) and others.

32.2.2 Default Information Systems

We generalize the theory of information systems by simply adding a default component. We should mention at this point that we limit ourselves to the so-called *normal* default structures. The reason for this is not that we cannot define general default rules, but rather that there are problems with the existence of extensions in the full case that we want to avoid.

Definition 3 A normal default information structure is a tuple

$$\underline{A} = (A,\ Con, \Delta,\ \vdash)$$

where $(A, Con, \vdash)$ is an information system, Δ is a set of pairs (X, Y) of consistent finite subsets of A, each element of which is written as $\dfrac{X : Y}{Y}$.

In our application, tokens will be "tuples" or *infons* of the form

$$\langle\!\langle \sigma, m_1, \ldots, m_n; i \rangle\!\rangle,$$

where σ is a relation name, the m_j are elements of a structure, and i is a "polarity" – either 0 or 1. The rules in Δ should therefore be read as follows. If the set of tuples X is in our current database, and if adding Y would not violate any database constraints, then add Y.

In default logic, the main concept is the idea of an extension. We define extensions in default model theory using Reiter's conditions, but extensions are now (partial) models. The following definition is just a reformulation, in information-theoretic terms, of Reiter's own notion of extension in default logic.

Definition 4 Let $\underline{A} = (A, \Delta, \vdash)$ be a default information structure, and x a member of $|\underline{A}|$. For any subset S, define $\Phi(x, S)$ to be the union $\bigcup_{i\in\omega} \phi(x, S, i)$, where

$$\begin{aligned}
\phi(x, S, 0) &= x, \\
\phi(x, S, i+1) &= \overline{\phi(x, S, i)} \cup \bigcup\{Y \mid \frac{X : Y}{Y} \in \Delta \;\&\; X \subseteq \phi(x, S, i) \;\& \\
&\qquad\qquad Y \cup S \in Con\}.
\end{aligned}$$

y is an extension of x if $\Phi(x, y) = y$. In this case we also write $x\delta_{\underline{A}}y$, with the subscript omitted from time to time.

Example: Default approximate reals. Use the information system described above. We might like to say that "by default, a real number is either between 0 and 1, or is the number π". We could express this by letting Δ consist of the rules $\frac{:Y}{Y}$, where Y ranges over singleton sets of rational pairs $\{\langle p,q\rangle\}$ such that $p \le 0$ and $q \ge 1$, together with those pairs $\{\langle r,s\rangle\}$ such that $r < \pi$ and $s > \pi$. Then, in the ideal domain, the only extension of $[-1,2]$ would be $[0,1]$; the interval $[-2,0.5]$ would have $[0,0.5]$ as an extension, and there would be 2 extensions of $[-2,4]$, namely $[0,1]$ and $[\pi,\pi]$.

In the full paper we show that all of this material can be stated in order-theoretic terms, without the need for information systems. This will make it possible to see the essential formula defining extensions, and will give a hint as to why we believe the order-theoretic approach is an interesting one to take.

32.3 Constraint Default Structures for First-order Logic

Assume, for purposes of this paper, that we are given a signature for first-order logic without equality, and with no function symbols other than constants. We will interpret first order logic using a nonstandard class of models. Our structures will be default information systems based on a particular set of individuals M. We first have to assume some *constraints* on any relations which are going to be holding in such sets M. These constraints will be used to generate the monotonic forcing relation $\vdash$ in the default structure. (The defaults themselves can be arbitrary, as long as they are normal.) We can use sets C of arbitrary closed formulas of first-order logic to state background constraints; in fact, we can use any language for which first-order structures are appropriate models.

To interpret formulas, we first of all choose some set M of individuals. We do *not* fix relations on M as in the standard first-order case, but we do choose particular individuals to interpret the constants[1]. Now, tokens will be infons of the form

$$\sigma = \langle\!\langle R, m_1, \ldots, m_n; i \rangle\!\rangle$$

where R is a relation name, $m_j \in M$, and $i \in \{0,1\}$. (This last item is called the *polarity* of the token.) We say that a set s of these tokens is *admissible* if (i) it does not contain any tokens conflicting in polarity, and (ii) it matches a model of C in the usual first-order sense. That is, there is a structure

$$\mathcal{M} = (M, (R_1, \ldots R_k))$$

[1] In terms of philosophy of language, we are taking constants to be rigid designators.

where the R_j are relations on M of the appropriate arities, such that $\mathcal{M}$ is a model of C, and such that

$$\langle\!\langle R_j, m_1, \ldots, m_n; 1\rangle\!\rangle \in s \Rightarrow R_j(m_1, \ldots, m_n) \text{ is true in } \mathcal{M}.$$

Similarly,

$$\langle\!\langle R_j, m_1, \ldots, m_n; 0\rangle\!\rangle \in s \Rightarrow R_j(m_1, \ldots, m_n) \text{ is false.}$$

An admissible set of infons is *total* if it is maximal in the subset ordering on sets of infons. A total set is isomorphic to an ordinary first-order structure $\mathcal{M}$.

Now we can specify a default information structure relative to M and C. Actually, the work is in specifying the strict (monotonic) part of the system. The defaults can be arbitrary normal ones.

Definition 5 Let M be a set, and C a constraint set. A first-order default information structure relative to M and C is a structure of the form

$$\underline{A}(M, C) = (\, A,\ \mathit{Con},\ \Delta,\ \vdash\,)$$

where A is the token set described above. A finite set X of tokens will be in *Con* if it is admissible, and $X \vdash \sigma$ iff for any total admissible set t, if $X \subseteq t$ then $\sigma \in t$.

Examples. The above definition encodes the constraints C into the $\vdash$ relation of the information system. For example, consider the constraint obtained by taking C to be the true formula $\mathbf{t}$. Intuitively, this should be no constraint at all, so our entailment relation should be the minimal one in which $X \vdash \sigma$ if and only if $\sigma \in X$. This is in fact the case. First notice that because $C = \mathbf{t}$, that a total admissible set t is one which (i) contains no infon $\sigma = \langle\!\langle \sigma, m; i\rangle\!\rangle$ and the *dual* infon $\overline{\sigma}$ of opposite polarity; and (ii) for any infon σ, contains either σ or $\overline{\sigma}$. Now let X be a finite set of infons. If $X \vdash \sigma$ then by properties of information systems, the dual infon $\overline{\sigma} \notin X$. By definition of $\vdash$, for any total admissible set t of infons, if $X \subseteq t$ then $\sigma \in t$. If σ is not in X, let t be a total admissible set containing X and the infon $\overline{\sigma}$ of opposite polarity. Then both σ and $\overline{\sigma}$ would be in t, which is not possible for an admissible set.

Notice that our general definition is easily modified to particular classes of interpretations. For example, our constraints may be stated for just one intended model, say the real numbers with addition and multiplication. In that case, we choose our sets M to be allowable by the particular interpretation class, and we change the definition of admissibility so that first-order structures are chosen from our particular class as well. Technically, we should restrict M to be countable, so that our Scott domain is in fact ω-algebraic. In fact, though, we will mostly be interested in *finite* default models for first-order logic.

32.3.1 Syntax and Semantics

For lack of space, we omit the official details of our three-valued semantics; but they are standard, given a knowledge of the strong Kleene truth tables.

32.3.2 Nonmonotonic Consequence

Our semantics can now be used to define a relation of nonmonotonic entailment, written $\mid\!\sim$, between sentences of our (positive) first-order logic. Understanding this notion is a step towards understanding the probabilistic measure introduced in the next subsection.

Intuitively, when we say that φ nonmonotonically entails ψ, we mean that having only the information φ, we can "leap to the conclusion" ψ. The usual example is that, knowing only that Tweety is a bird, we can leap to the conclusion that Tweety flies, even though penguins do not fly. A great deal of effort in the AI community has gone into giving a proper interpretation to the assertion $\varphi \mid\!\sim \psi$. We use (finite) default models and extensions to interpret it.

The notion of "only knowing" φ in $\varphi \mid\!\sim \psi$ (Levesque 1990), given a default information structure, is captured by interpreting the antecedent φ in a certain small class of situations for the structure. There are at least two possibilities for this class. One natural one is to use all set-theoretically minimal situations supporting φ. The second is to interpret φ in the *supremum closure* of these minimal models. We choose the second in this paper, because it seems better motivated from the probabilistic standpoint to be given in the next subsection.

We therefore make the following definitions.

Definition 6 Let $\underline{A}(M,C)$ be a default information structure, and φ a sentence of our logic. Let s,t range over situations.

- $MM(\varphi)$ is the set $\{s \mid s \text{ is minimal such that } s \models \varphi\}$;
- $U(\varphi)$ is the **supremum closure** of $MM(\varphi)$: the collection of situations obtained by taking consistent least upper bounds of arbitrary subcollections of $MM(\varphi)$. If $s \in U(\varphi)$ we will say that s is a minimal-closure model of φ.

Notice that since our logic is positive, every situation in $U(\varphi)$ will support φ.

Given these concepts, we can define nonmonotonic consequence (in a structure) as follows.

Definition 7 Let φ and β be sentences in first-order logic. Let $\underline{A} = A(M,C)$ be a finite normal default information system as above. We say that $\varphi \mid\!\sim_{\underline{A}} \beta$ if for all minimal-closure models $s \in U(\varphi)$,

$$\forall t: \ t \text{ is an } \underline{A}\text{-extension of } s \ \Rightarrow t \models \beta.$$

Example. We give the standard bird-penguin example. Assume that our language contains two predicates *Bird* and *Penguin*, and that *Tweety* is a constant. Let C be the constraint

$$(\forall x)(Penguin(x) \to Bird(x) \wedge \neg Fly(x)).$$

Consider a structure $A(M, C)$. Form the defaults

$$\frac{\langle\!\langle bird, m; 1\rangle\!\rangle : \langle\!\langle fly, m; 1\rangle\!\rangle}{\langle\!\langle fly, m; 1\rangle\!\rangle}$$

for each m in M. These defaults express the rule that birds normally fly. We then have

$$MM(Bird(Tweety)) = U(Bird(Tweety)) = \{\{\langle\!\langle Bird, tw; 1\rangle\!\rangle\}\}$$

where tw is the element of M interpreting *Tweety*.

The only extension of $\{\langle\!\langle Bird, tw, ; 1\rangle\!\rangle\}$ is $\{\langle\!\langle bird, tw; 1\rangle\!\rangle, \langle\!\langle fly, tw; 1\rangle\!\rangle\}$. Therefore

$$Bird(Tweety) \mid\!\sim Fly(Tweety).$$

We do not have $Penguin(Tweety) \mid\!\sim Fly(Tweety)$, because of the constraint C.

32.4 Probabilities, Constraints, and Attunement

Where do the default structures (in particular the default systems above) come from? We suggest that they could come from *default constraints.* Consider a (syntactic) construct of the form $\phi(x) \mid\!\sim \psi(x)$ where x is a free variable (perhaps *parameter*). Then, given a structure $A = A(M, C)$ an *admissible default system* would be one where the set of defaults Δ is such that with respect to all anchorings α of the free variables, we have that $(\phi \mid\!\sim_A \psi)[\alpha]$. By this last notation, we mean just to write out the definition of $\mid\!\sim$ again, but with respect to the anchoring α. A more stringent notion of consequence is now possible, as we can insist that in a structure, one formula entails another with respect to any admissible default system.

In fact, though, the usual default sets seem to come about in other ways. The example of the Nixon Diamond will serve to illustrate this point. In this example, Quakers are by default pacifists, and Republicans by default warmongers, and Nixon is strictly a Quaker and a Republican. The default sets satisfying the two default constraints are usually lumped together, with the result that one extension has Nixon as a warmonger, and another has him as a pacifist. No one so far has tried to separate default sets, constructing extensions in stages, to our knowledge.

The fact that default sets can be arbitrary has other amusing ramifications. We can use defaults to generate *degrees of belief* or *subjective probabilities* of various logical statements. By "subjective probability" we mean an analogue of the usual probability, a number that would be assigned to a statement by a particular agent or subject, given a default system and some basic constraints on the world. Let us illustrate with an example of Poole (Poole 1989.)

32.4.1 Poole's Anomaly

Assume that there are exactly three mutually exclusive types of birds: penguins, hummingbirds, and sandpipers. It is known that penguins don't fly, that hummingbirds are not big, and that sandpipers don't nest in trees. Now suppose we want to assert that the typical bird flies. Since we only speak of birds, we can do this with the precondition-free "open default constraint"

$$\mathbf{true} \mid\!\sim fly(x).$$

We would also like to say that the typical bird is (fairly) big, and that it nests in a tree. Similar defaults are constructed to express these beliefs.

The " paradox" is that it is impossible now to believe that Tweety, the typical bird, flies. To see why, let us formalize the problem more fully in our first-order language. Let C_1 be the obvious first-order sentence asserting that every individual is one of the three types of birds, and that no individual is of more than one type. Let C_2 be the conjunction of three sentences expressing abnormalities. One of these is, for example,

$$\forall x(\mathit{Penguin}(x) \to \neg \mathit{fly}(x)).$$

Let the background constraint C be $C_1 \wedge C_2$.

The defaults must be given in the semantics. Let M be a finite set, and consider the first-order default structure A with the admissible set of precondition-free defaults

$$\frac{:\langle\langle \mathit{fly}, m; 1\rangle\rangle}{\langle\langle \mathit{fly}, m; 1\rangle\rangle}; \frac{:\langle\langle \mathit{big}, m; 1\rangle\rangle}{\langle\langle \mathit{big}, m; 1\rangle\rangle}; \frac{:\langle\langle \mathit{tree}, m; 1\rangle\rangle}{\langle\langle \mathit{tree}, m; 1\rangle\rangle}.$$

We need only precondition-free defaults, because we only speak about birds. Further, we need no infons mentioning penguins, or any of the other species. The constraints can still operate.

We assert that if M has n elements, then there are 3^n extensions of the empty set (which is in fact the least model of $\mathbf{t}$). This is because any extension will include, for each bird $m \in M$, exactly two out of the three infons $\langle\langle fly, m; 1\rangle\rangle, \langle\langle big, m; 1\rangle\rangle, \langle\langle tree, m; 1\rangle\rangle$. The extension cannot contain all three infons, because the constraints rule that out. So each of n birds has three choices, leading to 3^n extensions.

One such extension is

$$\{\langle\langle \mathit{big}, m; 1\rangle\rangle : m \in M\} \cup \{\langle\langle \mathit{tree}, m; 1\rangle\rangle : m \in M\}$$

which omits any infons of the form $\langle\langle \mathit{fly}, m; 1\rangle\rangle$. This extension is one where *no* birds fly, where all birds are penguin-like. So now if *Tweety* is a constant of our language, then the formula *Fly(Tweety)* is not a nonmonotonic consequence of the "true" formula **true**, whose minimal model is the empty set. Further, if we move to the situation of seventeen bird types, each with its own distinguishing feature, we still have the case that Tweety cannot be believed to be flying. Poole suggests that this raises a problem for most default reasoning systems.

32.4.2 A Pseudo-probabilistic Solution

We now contend that the problem is not so severe. Notice that it is only in 3^{n-1} extensions that Tweety does not fly. This is because in an extension where Tweety does not fly, the constraints force the infons involving Tweety to assert that he is big and lives in a tree. Tweety thus only has one non-flying choice. The other $n-1$ birds have the same three choices as before. It seems therefore truthful to say that with probability $(3^n - 3^{n-1})/3^n = 2/3$, Tweety believably does fly. Moreover, imagine a scenario with seventeen mutually exclusive bird types, the same kinds of exceptions for each type, and defaults for all of the types. Then we would get that Tweety flies with probability 16/17.

We use this example to define our notion of subjective degree of belief:

Definition 8 Let φ be a sentence of positive first-order logic. Assume that relative to a constraint C, φ has minimal models in a structure M with n elements. Then the **conditional subjective degree of belief** $Pr([\psi \mid \varphi]; n, C)$ is defined to be the quantity

$$\frac{1}{N} \sum_{s \in U(\varphi)} \frac{card\{e : s \; \delta \; e \;\&\; e \models \psi\}}{card\{e : s \; \delta \; e\}}$$

where N is the cardinality of $U(\varphi)$.

Example. Referring to the case of Tweety above, we have

$$Pr([Fly(Tweety) \mid \mathbf{t}]; n, C) = 2/3.$$

Note the non-dependence on n. This raises a question about limits as n gets large, a topic which we must defer here. The full story is more nearly told in (Grove et al. 1992), which refers to the notion of *limit law* for models in first-order logic. The problem seems to be that in many examples of the above type, there is covergence of the subjective degree of belief measure in models of size n as n grows without bound. However, there seems to be no characterization of exactly when this happens, and simple (non-natural) examples show that limits need not always exist.

Here is what we mean now by "attunement." Notice that if we change our set of predicates and constraints to the case of seventeen bird types, but retain the rule system for three types, then we still get the same degree of belief (2/3) for Tweety's flying. Imagine that the universe had had seventeen bird types all along, with the constraints for those types. Then our agent, living in a small portion of the world (is there a situation with penguins and hummingbirds in it?) might have only observed the three types of birds. In that case, her 2/3 subjective degree of belief would not be as correct as it could be. Traveling to Australia might help refine the defaults.

Our definition bears a strong resemblance to the notion defined by Bacchus, Grove, Halpern, and Koller (Bacchus et al. 1993.) Their definition of the conditional probability of a statement ψ given another statement φ, though, is not made with reference to a given default information system. Instead, defaults are "translated" into a special logic for reasoning about statistical information. (For example, one can say that the proportion of flying birds out of all birds is approximately .9). Then, the translated default statements, and the given formulas φ and ψ are given a conditional probability in a standard first-order structure. Our corresponding "translation" of default statements is into a system of default rules, just as in Reiter's formulation. Our semantics also contrasts with that of BGHK in that it looks at partial worlds as well as total ones, and can assign degrees of belief to a statement's not being resolved one way or another.

32.4.2.1 Nixon Revisited

The method of default model theory can be adapted to differing kinds of logics for reasoning about default models. This will help us make use of a more specific logical language, should that be appropriate. We illustrate this with an improved model of the Nixon Diamond, using the BGHK probabilistic language.

Example (Nixon.) A fraction α of quakers are pacifists, and a fraction β of republicans are non-pacifists. These are our constraints on any actual world, but which people are pacifists currently is not known. Our version of the logic of BGHK has as atomic formulas

$$\rho x.\varphi(x) \approx_i \alpha,$$

which means that "the proportion of elements x satisfying $\varphi(x)$ is approximately α." Here α is a rational fraction in $[0, 1]$, and the subscript i refers to the i-th component of a "tolerance vector" of positive reals

$$\tau = \langle \tau_1, \ldots, \tau_i, \ldots \rangle$$

which is supplied with a standard finite first-order structure M. The semantics is that $(M, \tau) \models \rho x.\varphi(x) \approx_i \alpha$ if the fraction of domain elements satisfying φ is within τ_i of α. Here we set $i = 1$ and can actually fix $\tau_i = 0$. We thus want our background constraints to be

$$C(\alpha, \beta) = \rho x.(Pac(x) \mid Qu(x)) \approx \alpha \wedge \rho x.(\neg Pac(x) \mid Rep(x)) \approx \beta.$$

This formula uses *conditional expressions* of the form $\rho x.(\psi \mid \theta)$, the semantics of which in BGHK are a bit tricky when there are no domain elements satisfying θ, but which are not a problem in our case, as the expression denotes the fraction of domain elements satisfying $\psi \wedge \theta$ divided by the fraction satisfying θ, and we will always have positive numbers in the denominator.

We are interested in what happens as we vary α and β. But we keep these parameters fixed for what follows. Suppose now that our given information is "there are exactly N real quakers" and "there are exactly M real republicans.", and that there is exactly one quaker-republican, and that Nixon is that one. The Bacchus logic cannot easily express such statements. So instead of calculating a conditional "probability", we just consider a world which has exactly this information. We further simplify matters by assuming that N and M are chosen so that the numbers αN and βM are whole numbers. We consider a model of size $N+M-1$. What is the degree of belief in Nixon's being a definite, true pacifist? We assume our model consists of the integers from 1 to $N+M-1$ and interpret Nixon as N.

One world satisfying our conditions is a situation s_0 containing the infons

$$\langle\langle qu, n; 1\rangle\rangle \text{ for } 1 \leq n \leq N$$

and

$$\langle\langle rep, m; 1\rangle\rangle \text{ for } N \leq m \leq N+M-1.$$

Also we have N is Nixon. Now any permutation of the set of $N+M-1$ elements keeping Nixon fixed will do as another minimal situation expressing that there are exactly N quakers and M republicans. These would all assigned equal probability, and the computations would be the same in each case. Thus to get the subjective degree of belief of Nixon being a true pacifist, it suffices to consider the situation first described, and to calculate the fraction of extensions of this situation in which Nixon is actually a true pacifist. This is the degree to which $Pac(N)$ is believed, or the subjective probability of $Pac(N)$.

In this setting, we are interested in the probability of a formula's not being supported one way or another. So by $\neg\phi$ we will now mean the *weak* negation of ϕ. A situation will support $\neg\phi$ iff it does not support ϕ. We do want to talk about true warmongers, and we will do this with new predicate symbols. The predicate $NPac(x)$ will be interpreted as true warmongering. We therefore have to add the constraint that no individual is a true warmonger and a true pacifist at the same time.

Our infons will have the form $\langle\langle \sigma, m; i\rangle\rangle$, where σ is one of $\{rep, qu, pac\}$. (The predicate symbol $NPac$ will be interpreted by $\langle\langle pac, m; 0\rangle\rangle$.)

We could calculate degrees of belief (1) when there are no defaults, (2), when we have the default constraint only that typically republicans are warmongers, (3) when we have only the default constraint that typically quakers are pacifists, and (4) when we have both (2) and (3). The defaults satisfying these constraints are taken to be

$$\frac{\langle\langle rep, m; 1\rangle\rangle : \langle\langle pac, m; 0\rangle\rangle}{\langle\langle pac, m; 0\rangle\rangle} \; (N \leq m \leq N + M + 1);$$

and

$$\frac{\langle\langle qu, n; 1\rangle\rangle : \langle\langle pac, n; 1\rangle\rangle}{\langle\langle pac, n; 1\rangle\rangle} \; (1 \leq n \leq N).$$

The first case is easily handled. There is only one extension, namely the current world. In this world, true pacifism is not known, and true warmongering is not known. So "neither", namely $\neg Pac(N) \wedge \neg NPac(N)$ has probability 1. $Pac(N)$ and $NPac(N)$ both have probability 0.

We omit cases (2) and (3) for lack of space, and proceed to case (4). A detailed calculation reveals

$$Pr(Pac(N) \mid s_0) = \frac{\alpha(1-\beta)}{1-\alpha\beta}.$$

Similarly, the degree of belief in Nixon's warmongering is

$$Pr(NPac(N) \mid s_0) = \frac{\beta(1-\alpha)}{1-\alpha\beta}.$$

The degree of belief in neither (the agnostic position) is

$$Pr(\neg(Pac(N) \vee NPac(N)) \mid s_0) = \frac{1-(\alpha+\beta)+\alpha\beta}{1-\alpha\beta}.$$

In all of these expressions s_0 is just our starting situation.

Notice how the choice of α and β influences the number $\frac{\alpha(1-\beta)}{1-\alpha\beta}$ If they are chosen close to 1, the value is indeterminate, unless assumptions are made about how α and β approach 1. When $\alpha = \beta$, for example, then we get a .5 limit. We also get a .5 limit for warmongering, and an 0 limit for being undecided. On the other hand, when $\alpha = \beta = 0$, we get all of the weight on being undecided. But this is natural given that our defaults "program" us to assume that quakers are normally pacifists, and republicans are normally warmongers, in the face of the background constraint that there are no quaker pacifists, nor any republican warmongers. In this case we are stuck with our initial situation. Again we see a case of attunement. If our default constraints are unrealistic, we cannot use them to get information.

References

Bacchus, F., A. Grove, J. Halpern, and D. Koller. 1993. Statistical Foundations for Default Reasoning. In *Proceedings of IJCAI.*

Barwise, Jon. 1989. *The Situation in Logic.* 17. Stanford, California: Center for Study of Language and Information.

Cavedon, Lawrence. 1994. Constraints, Channels, Conditionals. This proceedings.

Devlin, K. J. 1991. *Logic and Information.* Cambridge University Press.

Etherington, D. W. 1988. *Reasoning with incomplete information.* Research Notes in Artificial Intelligence. Morgan Kaufman.

Grove, A., J. Halpern, and D. Koller. 1992. Asymptotic conditional probabilities for first-order logic. In *Proc. 24th ACM Symp. on Theory of Computing*, 294–305.

Jr., H. Kyburg. 1961. *Probability and the Logic of Rational Belief.* Wesleyan University Press.

Kraus, S., D. Lehmann, and M. Magidor. 1990. Nonmonotonic reasoning, preferential models, and cumulative logics. *Artificial Intelligence* 44:167–207.

Levesque, H. J. 1990. All I Know: A study in autoepistemic logic. *Artificial Intelligence* 42:263–309.

Marek, W., A. Nerode, and J. Remmel. 1990. A Theory of Nonmonotonic Rule Systems. In *Proceedings of 5th IEEE Symposium on Logic in Computer Science*, 79–94.

Poole, D. L. 1989. What the Lottery Paradox Tells us about Default Reasoning. In *Proceedings of First Annual Conference on Knowledge Representation.* Morgan Kaufmann.

Reiter, Raymond. 1980. A Logic for Default Reasoning. *Artificial Intelligence* 13:81–132.

Saraswat, V., M. Rinard, and P. Panangaden. 1991. Semantic foundations of concurrent constraint programming. In *Proceedings of POPL 1991.*

Scott, Dana S. 1982. Domains for Denotational Semantics. In *Lecture Notes in Computer Science 140.*

Seligman, J., and J.Barwise. 1993. Channel theory: towards a mathematics of imperfect information flow. June.

Zhang, Guo-Qiang. 1991. *Logic of Domains.* Boston: Birkhauser.

33

A Simultaneous Abstraction Calculus and Theories of Semantics

PETER RUHRBERG

Introduction

There are close similarities between some of the most influential recent semantic theories, such as Discourse Representation Theory (DRT), see Kamp 1981 and Zeevat 1991a, Dynamic Montague Grammar (DMG), see Groenendijk and Stokhof 1990, especially when based on the Dynamic Property Theory (DPT) of Chierchia 1994, and Situation Theory (ST), as described in Barwise and Cooper 1991, with its underlying theory of abstraction (AL) as it was developed in Aczel and Lunnon 1991. They all can be construed as involving forms of *simultaneous abstraction.* In DRT we have a set of discourse markers as one ingredient of a DRS, though one usually doesn't think of them as being abstracted over. We will show a pretty straightforward way of rendering them as abstractions. In DPT the $^{\sqcap}$-operator can be understood as abstracting over the infinite set of discourse markers, which are a subsort of the variables. In AL we take a one-one mapping of roles to parameters as input to abstract over a set of parameters in one go.

Abstraction over an unordered set of variables, or parameters, creates the problem of defining a sensible operation of application which unambiguously determines which of the objects applied to fill which of the positions abstracted over. The solution to this difficulty comes in the form of *application to assignments.* These abstracts denote, or correspond to, functions from assignments to objects of some kind. They are thus rather like open

I thank Lawrence Cavedon, Robin Cooper, Rachel Lunnon, and Paul Schweizer for criticism and support. This research was funded in part by ESPRIT BR6852 (DYANA-2) and LRE 62-051 (FraCaS).

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

formulas in standard logics which have their semantics given in terms of functions from assignments to denotations. In the same vein DRSs can be taken to denote, under an assignment, a set of verifying "embeddings," that is, a function from variable assignments to truth values. In DPT functions from discourse marker assignments to propositions are first class citizens of the semantic domain itself, and in AL the application operation is defined on assignments of entities to role indices.

It seems worthwhile to look for a common core which brings out the similarities as well as the differences between these theories more clearly and with formal precision. This core is our *Simultaneous Abstraction Calculus*. Semantic theories can be obtained from it by giving axioms of properties, propositions and truth, which again may be shared to some extent across competing theories. One may group the axioms into convenient packages under the banner of "abstract specification," thus creating a framework that should facilitate the comparison of existing theories as well as the development of alternatives. DMG and DRT for example, which we concentrate on in this paper, share a common theory of identity and propositional logic, but diverge in the kinds of quantification they employ. ST might be obtained by adding a logic of situation types to such unsituated theories of propositions. I can only give a sketchy account of these matters here, and also must leave the topic of relating these semantic objects to natural language utterances largely in the background.

Variables: "Bound yet Free"

As a consequence of the idea of simultaneous abstraction we will face a certain *loss of α-equivalence*. This is an inevitable result of the way in which application has to work for such terms. A renaming of variables will normally change the object as it will yield different results when applied to the same assignment. This should come as no surprise: it is of course well known that free variables cannot be renamed, but neither can the "abstracted" discourse markers of a DRS be changed without consequences. The same holds true for DPT discourse markers under $^\sqcap$-abstraction. Parameters in AL can be renamed, if abstracted over, but role indices cannot without changing the object.

I will not embark into an argument about whether it is wrong to speak of "abstraction" when there is no α-equivalence. It is more important to see that it is not a virtue in itself to be able to rename variables, as this means that computational work has to be done to establish a trivial identity of objects. Its virtue, as we shall see, lies in obtaining a total substitution operation, which is essential for the full exploitation of β-equivalences.

The idea that variables may be essential to the identity of a semantic object is certainly now fairly accepted, due to DRT and its "dynamic" children. From a more philosophical perspective I would argue that using variables in this way is no more a contamination of semantic objects by non semantic concepts than the common view that the order of arguments to a many place function should matter.

33.1 The Simultaneous Abstraction Calculus

In this section I introduce the basic formalism for simultaneous abstraction. I then add an operation of partial application to it, and compare the system to a similar one which is closer to Aczel and Lunnon's approach. Finally I take a look at possibly non well-founded structured objects.

33.1.1 The Basic Calculus

Instead of abstracting over single variables, as in standard λ-calculus, we allow λ-abstraction over any set of them. Terms are applied to records assigning terms to variables. For the rest of the paper, differently indexed variables are assumed to be non-identical.

Definition 1 The language Λ consists of TERMS $t, t' \ldots$ built up from basic type free CONSTANTS $c, c' \ldots$, and $\#$ for "undefined," and VARIABLES $x, y \ldots \in Var$ by means of ABSTRACTION $\lambda M.t$, for $M \subseteq Var$, and APPLICATION $t(x_1.t_1, \ldots, x_n.t_n)$.

We in fact allow infinite sets of variables to be abstracted over and also infinite records $(x_\imath.t_\imath)_{\imath \in I}$ to be applied to. We often write $\{x_1..x_n\}t$ instead of $\lambda\{x_1, .., x_n\}.t$, without meaning to imply that n has to be a finite number. A similar remark holds for application terms. Ways of introducing standard abstraction and application into such a system will be discussed later on.

Semantically, we will be working in the category CPO of complete partial orders and continuous functions. So a DOMAIN will always be understood to be a cpo with a least element $\perp$, a mapping between cpo's to be continuous, and the operation $\rightarrow$ to form the space of continuous functions, ordered pointwise.[1] g_f^M is defined by $g_f^M(x) = f(x)$ if $x \in M$, otherwise $g_f^M(x) = g(x)$, for all $x \in Var$. We write $\langle x_\imath \mapsto \xi_\imath \rangle_{\imath \in I}$ for the function f such that $f(x_\imath) = \xi_\imath$ for $i \in I$ and $f(x) = \perp$ otherwise. A RETRACTION between two domains, written $D \rightleftharpoons_\Phi^\Psi C$, is a pair of strict mappings $\Psi : D \rightarrow C$ and $\Phi : C \rightarrow D$ such that $\Psi \circ \Phi = id_C$.

[1] In some cases they will be required to be *strict*, preserving the bottom element. For background information notions see Barendregt 1984. Notice that every partial assignment function is continuous (if strict), given that Var is a flat domain with an added $\perp$.

Definition 2 A Λ-MODEL consists of a retraction $D \rightleftarrows_{\Phi}^{\Psi} (\mathit{Var} \to D) \to D$, and an interpretation $\Im$ which maps constants into D, with $\Im(\#) = \bot_D$. The denotation of terms under a variable assignment g is given by:[2]

- $\|c\|^g = \Im(c)$;
- $\|x\|^g = g(x)$;
- $\|\lambda M.t\|^g = \Phi\ \lambda f \|t\|^{g_f^M}$;
- $\|t(x_\imath.t_\imath)_{\imath \in I}\|^g = \Psi(\|t\|^g)\,(\langle x_\imath \mapsto \|t_\imath\|^g\rangle_{\imath \in I})$.

The denotations are only well-defined because the following holds.[3]

Theorem 1 *$\lambda f \|t\|^{g_f^M}$ is continuous for all terms t.*

Definition 3 The FREE VARIABLES of a term are defined by

- $FV(x) = \{x\}$;
- $FV(c) = \emptyset$;
- $FV(\lambda M.t) = FV(t) \setminus M$;
- $FV(t(x_\imath.t_\imath)_{\imath \in I}) = FV(t) \cup \bigcup_{\imath \in I} FV(t_\imath)$.

Lemma 2 *If $x \notin FV(t)$ then $\|t\|^g = \|t\|^{g_d^x}$ for all g and d.*

In such a case the denotation of t is independent of what we assign to x, but this does not mean one could rename x by some y of which t is independent as well. For example $\{x\}x$ is independent of x and y, but $\|\{x\}x\|^g \neq \|\{y\}y\|^g$. The latter becomes clear when we apply these two functions to suitable assignments e.g., $\|\{x\}x\,(x.a, y.b)\|^g = \|a\|^g \neq \|b\|^g = \|\{y\}y\,(x.a, y.b)\|^g$, presuming our model is non trivial.

Definition 4 SIMULTANEOUS SUBSTITUTION $[s_\jmath/y_\jmath]_{\jmath \in J}$ (dropping the index set when no confusion can arise) is a partially defined operation given by:

- $[s_\jmath/y_\jmath]_{\jmath \in J} x = s_\jmath$, if $x = y_\jmath$ for some $j \in J$, otherwise x;
- $[s_\jmath/y_\jmath]c = c$;
- $[s_\jmath/y_\jmath]_{\jmath \in J} \lambda M.t = \lambda M.[s_\imath/y_\imath]_{\imath \in I} t$, where $I = J \setminus \{j \mid y_\jmath \in M\}$, if $M \cap \bigcup_{\imath \in I} FV(s_\imath) = \emptyset$, otherwise undefined;
- $[s_\jmath/y_\jmath]\ t(x_\imath.t_\imath)_{\imath \in I} = [s_\jmath/y_\jmath]t\ (x_\imath.[s_\jmath/y_\jmath]t_\imath)_{\imath \in I}$.

An example of an undefined substitution would be $[y/x]\{y\}t_{xy}$, where replacing x in t_{xy} by y would give undesired results and renaming y is not a

[2]The third clause has its predecessor in Zeevat 1991b, where abstraction over the sets of variables of a common type was defined in a reconstruction of DMG. The 'λ' on the right is used as an expression of the meta language in this clause, with the expected meaning.

[3]Proofs are generally omitted from this paper. None of them is particularly interesting or involved.

possibility. To avoid such trouble we require a FRESH VARIABLE for a term to occur neither free nor bound in it.

Let us say that $\Lambda \models t = t'$ iff for every Λ-model and assignment $\|t\|^g = \|t'\|^g$.

Theorem 3 1. $\Lambda \not\models \{x_1..x_n\}t = \{y_1..y_n\}[y_i/x_i]t$,
2. $\Lambda \not\models \{x_1..x_n\}\{x_{n+1}..x_{n+m}\}t = \{x_1..x_{n+m}\}t$.

Lemma 4 *$\|t\|^\gamma = \|[t_i/x_i]_{i\in I}\, t\|^g$ if $[t_i/x_i]t$ is defined, where $\gamma = g \wr \langle x_i \mapsto \|t_i\|^g\rangle_{i\in I}$.*

The following equations can now be shown to hold, where the list is not meant to be complete. The fact that substitution is not always defined is responsible for the difficulties in giving such a list. Fact 5.3, which allows for some amount of renaming of variables, is just one example of an equation that cannot be derived from 5.2 for this reason. It seems a good idea to look for a complete calculus for Λ_{AL}, defined below, before doing the same for Λ since substitution can be totally defined in that system.

Theorem 5 1. $\Lambda \models \lambda M.t = \lambda N.t$ *for* $N = M \cap FV(t)$*;*
2. $\Lambda \models \{x_i\}_{i\in I}t\,(x_j.t_j)_{j\in J} = [t_j/x_j, \#/x_i]_{j\in J', i\in I'}\, t$,
with $J' = J \cap I$ *and* $I' = I \setminus J$, *if the substitution is defined.*
3. $\Lambda \models \{x_i\}_{i\in I}t\,(x_i.t_i)_{i\in I} = \{z_i\}_{i\in I}[z_i/x_i]_{i\in I}t\,(z_i.t_i)_{i\in I}$,
where all z_i *are fresh.*

33.1.2 Partial Application

It is very natural to ask for an operation for partially filling the argument roles of a relation, leaving some of them simply open for more to come. We use the notation $t[x_1.t_1, .., x_n.t_n]$ for partial application. The semantics is as follows:

$$\|t[x_i.t_i]_{i\in I}\|^g =_{df} \Phi\lambda f\ \Psi\|t\|^g\,(f^{x_i|i\in I}_{\langle x_i\mapsto\|t_i\|^g\rangle_{i\in I}})\ .$$

The following properties of partial application are of particular interest, letting application associate to the left.

Theorem 6 1. $\Lambda \models \{x_i\}_{i\in I}t\,[x_j.t_j]_{j\in J} = [t_j/x_j]_{j\in J'}\{x_i\}_{i\in I'}t$,
if defined, where $J' = J \cap I$ *and* $I' = I \setminus J$*;*
2. $\Lambda \models t[x_i.t_i]_{i\in I}[y_j.s_j]_{j\in J} = t[x_i.t_i, y_j.s_j]_{i\in I, j\in J'}$
where $J' = J \setminus \{j \mid \exists i\ y_j = x_i\}$*;*
3. $\Lambda \models t[x_i.t_i]_{i\in I}(y_j.s_j)_{j\in J} = t(x_i.t_i, y_j.s_j)_{i\in I, j\in J'}$
where $J' = J \setminus \{j \mid \exists i\ y_j = x_i\}$.

33.1.3 Aczel–Lunnon Abstraction

The SAC is closely related to the Extended Kamp Notation (EKN) of Barwise and Cooper 1991, which is based on Aczel and Lunnon's theory of abstraction. In EKN λ-abstraction operates on injective functions from a set of *role indices* to *parameters*. In our case those two notions are merged into one, the variables (thought of as argument roles), which allows us to replace the injections by simple sets. This makes sense because in contrast to AL our semantics does not treat the application to variables in a term like $t(x.y)$ as a filling of the role x by some "indeterminate object" y, but rather as a linking of two roles. There is a price we pay for our simple mindedness: we cannot link a role y with any role x inside the scope of an abstraction over y by a simple application. The same problem was encountered in undefined substitutions such as $[y/x]\{y\}t_{xy}$, which indicated a certain deficit of elegance for our system. The following system revises the basic SAC to overcome this problem, using AL-style syntax for abstraction. We will show then that no additional expressive power is gained from such a move.

Definition 5 The TERMS of Λ_{AL} are built up from CONSTANTS $c, \#, ..$ and PARAMETERS $X, Y, ..$ by means of ABSTRACTION $\lambda\chi.t$ over injective functions χ from parameters to variables (=roles),[4] and TOTAL/PARTIAL APPLICATION $t(x_\imath.t_\imath)_{\imath\in I}$ and $t[x_\imath.t_\imath]_{\imath\in I}$.

Semantically we stay in the same space of domains $D \rightleftharpoons^{\Psi}_{\Phi} (Var \rightarrow D) \rightarrow D$ as for Λ, with the only difference being that we now interpret under assignments of objects to parameters, called ANCHORS, and that these have to be obtained from assignments to roles when we define the meaning of abstraction.

Definition 6 A Λ_{AL}-MODEL is a Λ-model, where the denotation of terms under an anchor g is given by:

- $\|c\|^g = \Im(c)$;
- $\|X\|^g = g(X)$;
- $\|\lambda\chi.t\|^g = \Phi\ \lambda f \|t\|^{g \wr (f \circ \chi)}$;
- $\|t(x_\imath.t_\imath)_{\imath\in I}\|^g = \Psi(\|t\|^g)(\langle x_\imath \mapsto \|t_\imath\|^g\rangle_{\imath\in I})$.

This solves the problems described above as substitution is total by means of renaming parameters e.g., $[X/Y]\lambda\langle X \mapsto z\rangle.t_{XY} = \lambda\langle Z \mapsto z\rangle.t_{ZX}$. The question that needs to be addressed is whether there is more to be gained, namely terms which have no equivalent in the original language Λ. To show that there are no such terms we only face one difficulty, namely that of dealing with multiple uses of roles in different levels of abstraction.

[4]In AL, the functions take roles to parameters injectively. In turning them around we could be more liberal in allowing several parameters to be mapped to the same role, which is equivalent to renaming those parameters by one of them which then gets mapped to the role in question. Notice that associating one parameter with many roles in one abstraction clearly makes no sense. Our semantics for Λ_{AL} will not presuppose injectivity, but we will assume it here nevertheless for convenience.

Such "role recycling" is made easy by using different parameters in a term like $\lambda\langle Y \mapsto x\rangle\lambda\langle Z \mapsto x\rangle.t_{YZ}$. In creating a Λ-term with the same role x used twice we face the potential problem of the outer abstraction to become vacuous. The solution is to inject an intermediate role y to arrive at $\{x\}(\{y\}\{x\}t_{yx}\ (y.x))$. To have such roles available we assume some form of restriction on Λ_{AL} abstraction to that effect. Alternatively one might employ a way of expanding *Var* suitably for the purposes of translation. I will be sloppy about the details of this in the following elaboration.

Definition 7 The TRANSLATION $[x_\imath/X_\imath]^\dagger_{\imath\in I} : \Lambda_{AL} \to \Lambda$ under a substitution of all free parameters of a Λ_{AL}-term by variables is defined by:

- $[x_\imath/X_\imath]^\dagger_{\imath\in I} X_\imath = x_\imath$;
- $[x_\imath/X_\imath]^\dagger c = c$;
- $[x_\imath/X_\imath]^\dagger_{\imath\in I}\ \lambda\langle Y_\jmath \mapsto y_\jmath\rangle_{\jmath\in J}.t =$ $(\lambda\{z_\imath\}_{\imath\in I''}\lambda\{y_\jmath\}_{\jmath\in J}.[y_\jmath/Y_\jmath]_{\jmath\in J}[x_\imath/X_\imath]_{\imath\in I'}[z_\imath/X_\imath]_{\imath\in I''}{}^\dagger t)\,(z_\imath.x_\imath)_{\imath\in I''}$, where $K = I \setminus \{\imath|\exists\jmath_{\in J}\ X_\imath = Y_\jmath\}$, $I' = K \setminus \{\imath|\exists\jmath_{\in J}\ x_\imath = y_\jmath\}$, $I'' = K \cap \{\imath|\exists\jmath_{\in J}\ x_\imath = y_\jmath\}$ and the $z_\imath$ are fresh;
- $[x_\imath/X_\imath]^\dagger t(y_\jmath.t_\jmath)_{\jmath\in J} = [x_\imath/X_\imath]^\dagger t\ (y_\jmath.[x_\imath/X_\imath]^\dagger t_\jmath)_{\jmath\in J}$.

Lemma 7 $\|t\|^g = \|[x_\imath/X_\imath]^\dagger_{\imath\in I} t\|^{g\circ\pi}$, *where* $\pi = \langle x_\imath \mapsto X_\imath\rangle_{\imath\in I}$.

Translation from Λ to Λ_{AL} is a trivial matter of renaming variables by parameters, and abstracted sets by mappings, under some fixed bijection between Par and Var. Hence we conclude the following.

Theorem 8 *There are faithful translations from* Λ_{AL} *to* Λ *and vice versa.*

It is clear that we can define unary abstraction and application in Λ_{AL} in terms of a designated role used solely for that purpose. By the above theorem we see that the same can be done in Λ as well. The SAC thus contains the unary λ-calculus. In particular, α-equivalence for λ-terms becomes an instance of Theorem 5.3 above.

Corollary 9 λ *can be embedded into* Λ.

33.1.4 Structured Objects and Systems of Equations

Using some encoding of standard abstraction, which exists for Λ by the corollary, we define the following PREDICATION OPERATION:

$$\langle\!\langle t; (x_\imath.t_\imath)_{\imath\in I}\rangle\!\rangle =_{df} \lambda x.\,(x(t)\,(x_\imath.t_\imath)_{\imath\in I}).$$

Theorem 10 $\langle\!\langle t; (x_\imath.t_\imath)_{\imath\in I}\rangle\!\rangle = \langle\!\langle t'; (x_\imath.t'_\imath)_{\imath\in I}\rangle\!\rangle \to t = t' \wedge \bigwedge_{\imath\in I} t_\imath = t'_\imath$.

We thus have structured objects and operations of abstraction and application in place that appear to satisfy much of the needs of current Situation Theory. A further demand is to have non well-founded objects, obtained from cyclic sets of equations. To solve systems of equations[5] $(x_\imath = t_\imath)_{\imath \in I}$ we need to encode assignments by some kind of terms. We could use predication for this, but it is still simpler to define $\ll x_\imath . t_\imath \gg_{\imath \in I} =_{df} \lambda x . x(x_\imath . t_\imath)_{\imath \in I}$ and $\pi_z =_{df} \lambda x . x(\{z\} z)$. Notice that

$$\pi_{x_\imath}(\ll x_\jmath . t_\jmath \gg_{\jmath \in J}) = t_\imath \quad for \ \ \imath \in J.$$

We say that an assignment $\bar{a}$ SOLVES A SYSTEM OF EQUATIONS $(x_\imath = t_\imath)_{\imath \in I}$ iff $\pi_{x_\imath}(\bar{a}) = [\pi_{x_\jmath}(\bar{a})/x_\jmath]_{\jmath \in I} t_\imath$ for $i \in I$.

Theorem 11 *Every system of equations has a solution.*

Proof. Recall that $\Upsilon =_{df} \lambda f. (\lambda x. f(xx))(\lambda x. f(xx))$ is a fixed point combinator with $t(\Upsilon t) = (\Upsilon t)$ for any term t. Define from a system of equations $(x_\imath = t_\imath)_{\imath \in I}$ the functor $F =_{df} \lambda z. \ll x_\imath . [\pi_{x_\jmath}(z)/x_\jmath]_{\jmath \in I} t_\imath \gg_{\imath \in I}$. Then (ΥF) solves the system, since $\pi_{x_\imath}(\Upsilon F) = \pi_{x_\imath}(F(\Upsilon F)) = \pi_{x_\imath} \ll x_\imath . [\pi_{x_\jmath}(\Upsilon F)/x_\jmath]_{\jmath \in I} t_\imath \gg_{\imath \in I} = [\pi_{x_\jmath}(\Upsilon F)/x_\jmath]_{\jmath \in I} t_\imath$. □

As things stand we do not necessarily have unique solutions to such equations. It may be consistent to assume uniqueness at least for some kinds of systems, in the spirit of Aczel and Lunnon 1991.

33.1.5 Equations for Λ_{AL}

The complications in clause three of the translation from Λ_{AL} to Λ indicate why a complete and reasonably simple set of equations for Λ will be hard to find. The slightly more complex language Λ_{AL} seems more suitable for this task. So I give a set of equations for it that I hope to be complete if used in conjunction with rules for congruence relations. The notions of free parameters and substitution are assumed to be defined in the standard way.

Theorem 12 1. $\models t(x_\imath . t_\imath, x_\jmath . \#)_{\imath \in I, \jmath \in J} = t(x_\imath . t_\imath)_{\imath \in I}$;

2. $\models \lambda \chi . t = \lambda \xi . t$ *for* $\xi = \chi|_{FP(t)}$;
3. $\models \lambda \langle X_\imath \mapsto x_\imath \rangle_{\imath \in I} t = \lambda \langle Y_\imath \mapsto x_\imath \rangle_{\imath \in I} [Y_\imath / X_\imath]_{\imath \in I} t$ *for fresh* $Y_\imath$*'s;*
4. $\models \lambda \langle X_\imath \mapsto x_\imath \rangle_{\imath \in I} t \, (x_\jmath . t_\jmath)_{\jmath \in J} = [t_\jmath / X_\jmath, \# / X_\imath]_{\jmath \in J', \imath \in I'} t$ *with* $J' = J \cap I$ *and* $I' = I \setminus J$;
5. $\models \#(x_\imath . t_\imath)_\imath \in I = \#$;
6. $\models \lambda \chi . \# = \#$.

33.2 Dynamic Montague Grammar

I now show how the SAC can provide a framework for specifying semantic theories, especially recent ones, which do not easily fit into the traditional λ-calculus.

[5] We need to assume that at least three variables do not occur in such a system, to allow us to form the terms below without accidentally binding some of the $x_\imath$ inside them.

The basic idea for semantics as conceived of in this paper is to think of properties as functions from assignments to propositions.[6] I will use the term *ω-properties* to distinguish them from ordinary functions which take entities into propositions. Equipped with this notion, we can reconstruct DMG along the lines of Chierchia 1994.

33.2.1 Using the SAC for DMG

For DMG, variables need to come in two kinds: the DISCOURSE MARKERS DM, and the remaining ones, which I call META VARIABLES MV. I use greek letters $\alpha, \beta \ldots$ for meta variables, dotted letters $\dot{x}, \dot{y} \ldots$ for discourse markers, and $x, y \ldots$ for variables that can be of either kind. The sorting of our variables is not a matter of the kinds of denotation that they take. It is rather a matter of protecting some of them, namely the meta variables, from the influence of abstraction so that we can bind variables inside the scope of $^{\sqcap}$.

The crucial point of DMG can now be captured easily by abstracting and applying to the infinite set of discourse markers via the following definitions:

- $^{\sqcap}t =_{df} \lambda DM.t$, and
- $^{\sqcup}t =_{df} t(\dot{x}.\dot{x})_{\dot{x} \in DM}$.

All other abstractions and applications will be the traditional unary ones. Notice that every discourse marker is bound in a term of the form $^{\sqcap}t$, and occurs free in $^{\sqcup}t$. We thus can β-convert a term $\lambda\alpha.t(^{\sqcap}t')$ even if some discourse markers of t' end up in the scope of an abstraction in t. Discourse markers are not free in $^{\sqcap}t'$ and hence do not become bound by a new operator inside t. Another reduction we use in DMG beside the more standard β-conversions is $^{\sqcup\sqcap}t = t$.

The sentence 'A man walks in' can now be rendered in our notation as

$$\lambda\alpha.\Sigma\dot{x}(man(\dot{x}) \sqcap walkin(\dot{x}) \sqcap {}^{\sqcup}\alpha)$$

which can applied to $^{\sqcap}whistle(\dot{x})$ ('he whistles') to yield

$$\Sigma\dot{x}(man(\dot{x}) \sqcap walkin(\dot{x}) \sqcap whistle(\dot{x}))$$

The important action happens at the level of conjunction of ω-properties, which is achieved by abstracting over "possible continuations" of the discourse.[7]

[6] Propositions can be very fine grained—there is no need to think of propositions as sets of possible worlds or situations. Even so, Bealer's fondalee/rajneesh puzzle, see Bealer 1989, suggests that propositional functions cannot cope with all problems of intensional grain, or at least that function application should be distinguished from a structure preserving predication operation, as was done in 2.4.

[7] Lambda conversion on $^{\sqcap}whistle(\dot{x})$ is only one way to make the conjunction happen; Unification of $whistle(\dot{x})$ with the free variable α, as in UCG, is just as intuitive for this. Hence the choice of the term 'meta variables', reminiscent of its use in Zeevat 1991a.

We use unary abstraction and application in the system in order to get a fairly standard logic that does not require us to rethink the setup of DMG. We saw that these notions can be defined within the confines of Λ.[8] Logical operations are treated as unary or binary constants whose semantic behavior is captured by our theories of truth, which we choose to give separate from the general model theory in an axiomatic way. The following logical constants are singled out: $\cap, \cup, -, \supset, \Sigma, \Pi, \doteq$, intended to be the operations of conjunction, disjunction, negation, implication, existential and universal quantification, and equality. I will take $\cap, -, \Sigma, \doteq$ as primitive, and the others to be defined in the standard way.[9] Logical combinators are written in the conventional form, $t \cap t'$ for $\cap(t)(t')$, $\Sigma x\, t$ for $\Sigma(\lambda x.t)$ etc., to enhance readability.

33.2.2 Frege Structure Axioms

For semantics, we need a theory of truth for the propositions that we hope to denote by the Λ-terms. To that end, we introduce a formal language in which to state such a theory. One might do the job in an informal model theoretic way along the lines of Aczel 1980 but I will make things look more like the formal treatment of Turner 1990.

Definition 8 We define WFFs of FOL_Λ in the standard way:

- if $t, t' \in TERM$ then $T(t), t = t' \in WFF$,
- if $\psi, \phi \in WFF$ then $\neg\phi, \phi \wedge \psi, \exists\alpha\phi \in WFF$.

The interpretation of WFFs proceeds in the standard way. T is a truth predicate. We define $F(t) =_{df} T(-t)$ (falsity), $P(t) =_{df} T(t) \vee F(t)$ (being a proposition), and $PTY^n(t) =_{df} \forall x_1 \dots x_n\ (\bigwedge_{i=1..n} x_i \neq \#) \rightarrow P(t(x_1)\dots(x_n))$ (being an n-place property). Here are the basic axioms:

$$\begin{array}{rrl} & P(t \doteq t') \wedge & T(t \doteq t') \leftrightarrow t = t' \\ P(t) \rightarrow & P(-t) \wedge & T(-t) \leftrightarrow \neg T(t) \\ P(t) \wedge P(t') \rightarrow & P(t \cap t') \wedge & T(t \cap t') \leftrightarrow T(t) \wedge T(t') \\ PTY^1(t) \rightarrow & P(\Sigma\, t) \wedge & T(\Sigma t) \leftrightarrow \exists z \neq \#\, T(t(z)) \end{array}$$

33.2.3 Some DMG Translations

I only give a short sketch here and refer the reader to Chierchia 1994 and Chierchia 1992 for a more complete treatment. Sentence denotations in

[8] As an alternative to taking these unary operations as defined in terms of Λ-expressions, one could introduce them as basic operations, and give them a semantics in an expanded domain $D \rightleftharpoons^{\Psi}_{\Phi} ((Var \rightarrow D) \rightarrow D) + (D \rightarrow D)$, in the way Chierchia 1994 does.

[9] There may be reasons not to do this, such as to avoid unintended identities in attitude contexts, or to be able to use the less symmetric notion of implication of Aczel 1980.

DMG are not propositions, but *context change potentials (ccp's)*, where a ccp is a function from ω-properties to propositions. Two operations are handy to convert props into ccp's and back: $\uparrow t =_{df} \lambda\alpha(t \cap {}^{\sqcup}\alpha)$ and $\downarrow t =_{df} t({}^{\sqcap}true)$. Ccp's allow for clever ways of leaving holes within quantified structures to be filled by later material in the discourse so that one can use functional composition $t; t' =_{df} \lambda\alpha.t({}^{\sqcap}(t'(\alpha)))$ for discourse sequencing:

$$\text{`}{}^{\dot{x}}[\text{a}]\text{'} \Longrightarrow$$
$$\lambda\alpha\lambda\beta\, \lambda\gamma\Sigma\dot{x}({}^{\sqcup}\alpha\, \dot{x})\,{}^{\sqcap}({}^{\sqcup}\beta\,(\dot{x})(\gamma))$$

$$\text{`}{}^{x}[\text{a man}] \text{ walks in. } {}^{\dot{x}}[\text{he}] \text{ whistles.'} \Longrightarrow$$
$$[\lambda\gamma\Sigma\dot{x}(man(\dot{x}) \cap walkin(\dot{x}) \cap {}^{\sqcup}\gamma)]\,;\, [\lambda\gamma.\, whistle(\dot{x}) \cap {}^{\sqcup}\gamma]$$
$$= \lambda\gamma(\Sigma\dot{x}\; man(\dot{x}) \cap walkin(\dot{x}) \cap whistle(\dot{x}) \cap {}^{\sqcup}\gamma)$$

The flexibility of ccp's is shown by various possibilities to define dynamic kinds of universal quantification for donkey sentences:

$$\text{`every'} \Longrightarrow$$
$$\lambda\alpha\lambda\beta\uparrow\Pi\dot{x} - [({}^{\sqcup}\alpha\,\dot{x})({}^{\sqcap}- \downarrow({}^{\sqcup}\beta\,\dot{x}))]$$

$$\text{`man with } {}^{y}[\text{a donkey}]\text{'} \Longrightarrow$$
$$\lambda\dot{x}\;\lambda\gamma(man(\dot{x}) \cap \Sigma\dot{y}(donkey(\dot{y}) \cap with(\dot{x},\dot{y}) \cap {}^{\sqcup}\gamma))$$

$$\text{`beats } {}^{y}[\text{it}]\text{'} \Longrightarrow$$
$$\lambda\dot{x}\uparrow beat(\dot{x},\dot{y})$$

By intensional function application $t({}^{\sqcap}t')$ on the translations one gets:

$$\uparrow\Pi\dot{x} - [man(\dot{x}) \cap \Sigma\dot{y}(donkey(\dot{y}) \cap with(\dot{x},\dot{y}) \cap -(beat(\dot{x},\dot{y}) \cap true))]$$

An alternative to the above "strong" reading of the universal, would assign the translation:

$$\text{`every'} \Longrightarrow$$
$$\lambda\alpha\lambda\beta\uparrow\Pi\dot{x}[- \downarrow({}^{\sqcup}\alpha\,\dot{x}) \;\cup\; ({}^{\sqcup}\alpha\,\dot{x})({}^{\sqcap}\downarrow({}^{\sqcup}\beta\,\dot{x}))]$$

resulting in the "weak" reading:

$$\uparrow\Pi\dot{x}[-(man(\dot{x}) \cap \Sigma\dot{y}(donkey(\dot{y}) \cap with(\dot{x},\dot{y}) \cap true))$$
$$\cup(man(\dot{x}) \cap \Sigma\dot{y}(donkey(\dot{y}) \cap with(\dot{x},\dot{y}) \cap beat(\dot{x},\dot{y}) \cap true))]$$

33.3 Discourse Representation Theory

The SAC mirrors very closely the use of discourse referents in DRSs. The similarity with Zeevat's semantics for those, in terms of pairs consisting of a set of variables and a function from assignments to truth, is undeniable, see Zeevat 1991a. We have put propositions in the place of truth values and defined an abstraction operation that takes such pairs into appropriate new functions. In addition, we can iterate such abstractions indefinitely and we have a simple notion of application at our disposal.

Our approach is most closely related to the Situation Theoretic DRT of Cooper 1993 who uses EKN with Aczel–Lunnon abstraction to similar effect. As in Cooper's treatment our version of DRT imposes a need for some additional abstractions at the right level. I won't go into the resulting complications for the syntax–semantics interface here.

33.3.1 ω-Properties

To be a propositional function in the argument roles $x_1 \ldots x_n$ is defined by:[10]

$$PF^{x_1..x_n}(t) =_{df} \forall z_1..z_n \bigwedge_{i=1..n} z_i \neq \# \rightarrow P(t(x_1.z_1, \ldots, x_n.z_n)).$$

This is not a sufficiently strict notion of ω-properties to support the forthcoming definitions and theorems. How strict a notion of ω-property one ultimately wants remains to be seen, but a function f that for example maps an assignment $\langle x \mapsto d \rangle$ to the proposition p while mapping $\langle x' \mapsto d' \rangle$ to $-p$ seems useless for semantics. We exclude such pathological cases by requiring something to the effect that

$$PTY^M(d) \Rightarrow \forall N \; PF^N(d) \leftrightarrow M \subseteq N.$$

A too strict definition would have $=$ instead of $\subseteq$. This would exclude any abstracts from denoting ω-properties, because they are insensitive to overdefined assignments:

$$\Lambda \models \{x_1..x_n\}t\,(x_1.t_1, .., x_n.t_n) = \{x_1..x_n\}t\,(x_1.t_1, .., x_{n+m}.t_{n+m})$$

Still one may ask whether the application of a ω-property to an assignment should ever result in something other than a proposition or $\perp$. The only sensible possibility seems to be to have application to "underdefined" assignments result in a new ω-property whose argument roles are those that haven't been filled on the first attempt. We have introduced the operation of *partial application* for these purposes, which will also enable us to de-

[10] We will use an infinitary version of FOL_Λ to express the theory, though one might get away with axiom shemata in the finitary logic if we stick to finite n here. In particular the definition of $PTY^M(t)$ below would have to be replaced by an infinite number of implicational axioms.

fine the duplex conditions of modern DRT. Here is a farly weak but useful notion of ω-property.

Definition 9 $PTY^{\{x_\imath|\imath\in I\}}(t) =_{df} \bigwedge_{I\subseteq J} PF^{\{x_\imath|\imath\in J\}}(t) \wedge \bigwedge_{I\not\subseteq J} \neg PF^{\{x_\imath|\imath\in J\}}(t)$.

33.3.2 Kamp Structures

Let us now try to give a theory of ω-properties and propositions and truth that is capable of dealing with basic ideas of DRT. We add to the logical operations $\cap, -, \doteq$ two new quantificational operators, namely a non selective existential Ξ and a binary conditional $\gg$. All these are required to be (bi)strict: yielding $\bot$ whenever one of their arguments is $\bot$. Here are the axioms for Kamp structures:

$$\begin{array}{rrl} & P(t \doteq t') \wedge & T(t \doteq t') \leftrightarrow t = t' \\ P(t) \rightarrow & P(-t) \wedge & T(-t) \leftrightarrow \neg T(t) \\ P(t) \wedge P(t') \rightarrow & P(t \cap t') \wedge & T(t \cap t') \leftrightarrow T(t) \wedge T(t') \end{array}$$

$$PTY^{x_1..x_n}(t) \rightarrow P(\Xi t) \wedge T(\Xi t) \leftrightarrow \exists z_1..z_n \neq \#T(t(x_1.z_1, .., x_n.z_n))$$

$$PTY^{\{x_\imath|\imath\in I\}}(t) \wedge PTY^{\{x_\jmath|\jmath\in J\}}(s) \rightarrow P(t \gg s) \wedge$$

$$T(t \gg s) \leftrightarrow \overset{\imath\in I}{\forall z_\imath}\, T(\,t(x_\imath.z_\imath)_{\imath\in I}) \rightarrow \overset{\jmath\in J\setminus I}{\exists z_\jmath} \neq \#T(\,s(x_\imath.z_\imath, x_\jmath.z_\jmath)_{\imath\in J\cap I, \jmath\in J\setminus I})$$

Notice that only bound variables in s are quantified in the conditional. To get some kind of "dynamic conjunction" of DRSs we can introduce the operation $\oplus$ into the system.[11] For example by means of this definition:

$$\|t \oplus t'\|^g =_{df} \Phi\ \lambda f\ \Psi\|t\|^g(f) \cap \Psi\|t'\|^g(f)$$

The following consequences can now be derived:

Theorem 13 $\models \lambda M.t \oplus \lambda N.s = \lambda M \cup N\,.\,t \cap s$, *if* $M \cap FV(\lambda N.s) = N \cap FV(\lambda M.t) = \emptyset$.

Using partial application we can define DRT-type quantifiers Q_x for the so called "duplex conditions," by means of standard generalized quantifiers Q as relations between ω-properties. The weak and strong readings are obtained as follows:

$$t\langle Q_x\rangle^w t' =_{df} Q(\lambda z.\Xi\, t[x.z]\,, \lambda z.\Xi\, t \oplus t'[x.z]),$$

$$t\langle Q_x\rangle^s t' =_{df} Q(\lambda z.\Xi\, t[x.z]\,, \lambda z.\, t[x.z] \gg t'[x.z]).$$

33.3.3 True Dynamics

It is time now to compare the SAC and with systems like the Dynamic Predicate Logic (DPL) of Groenendijk and Stokhof 1991, that are based

[11]This operation is used in Cooper 1993 for the interpretation of discourse. It becomes definable in a slightly richer language that employs variables over assignments.

on interpretations in terms of relations between assignments. DPL replaces the non-standard DRT syntax by FOL formulas, but they have shown how to give such a relational interpretation for DRSs too. Let us view relations between assignments as functions from assignments to sets of assignments. Call sets of assignments ω-sets, and take conditions to denote just ω-sets. We then get the following denotation for a DRS:

$$\|\{x_1...x_n\}\phi\|^g_{DPL} = \{h \mid g \overset{\{x_1..x_n\}}{\sim} h \ \wedge \ h \in \|\phi\|\}.$$

This should be compared with extensionalized SAC denotations, which we obtain when we replace propositions by truth values. We then also get ω-sets as denotations under an assignment, but they are different from the DPL ones:

$$\|\{x_1...x_n\}\phi\|^g_{SAC} = \{h \mid g^{\{x_1..x_n\}}_{h} \in \|\phi\|\}.$$

The point is that the h's in our set can assign anything they like to the variables outside $\{x_1...x_n\}$, while for the DPL style of denotation they have to agree with the incoming g.

A natural development, carried further in Dekker 1993, is to view "information states" as such ω-sets, and then lift denotations of DRSs to functions from ω-sets to ω-sets, such as:

$$\|\phi\|_{DPL\uparrow}(G) = \bigcup_{g \in G} \|\phi\|^g_{DPL}$$

which is quite different from the trivial lifting from ϕ to $\lambda p.(p \oplus \phi)$ that might be used in our system.

The lack of a requirement for assignments to agree on certain variables has consequences for semantics. Thus, in contrast to DPL and most forms of DRT, we must abstract over variables in order to link them anaphorically to preceding discourse by $\oplus$. If we would use DPL style composition for discourse conjunction instead the link could not be sustained over more than two sentences. A similar divergence can be found in the treatment of the conditional. Again we have to put anaphoric variables into to abstracted set of the consequence, but only those which are anaphoric to the antecedent, not those which refer further back. Abstraction in Λ really is just a way of partitioning roles into different levels, in view of how we want to fill them or quantify over them.

There is something of a philosophical gap between the approach of this paper and the "truly dynamic" ones, which is brought out by these considerations. On the SAC view the functions from assignments to ω-properties are not regarded as ways of "updating" information states by incoming material. It makes no sense on this view to talk of "input" and "output" in reference to the assignments. The guiding intuition is rather that semantic objects should be plausible as things to have attitudes towards. These objects, I take it, are rather static creatures whose role in the changes of mental states are another matter.

33.4 Conclusion

We have defined a calculus of simultaneous abstraction and shown it to be equivalent to a less simple minded one that is more reminiscent of Aczel and Lunnon's system. We established the existence of structured, and non well-founded objects in the system, by an embedding of the lambda calculus in it. We then obtained versions of Dynamic Montague Grammar and Discourse Representation Theory, using axiomatic theories of truth. It seems plausible that the same can be done for Situation Theory as well, supporting the claim that we have found a framework which is general enough for a wide range of semantic theories which take notions of propositions and truth as fundamental for the enterprise. It is at odds though with forms of "dynamic" semantics that take the idea of updating (sets of) assignments as their starting point. Many important logical questions remained unanswered in this paper, but I hope to have shown that they may be worth asking.

References

Aczel, Peter. 1980. Frege Structures and the Notions of Proposition, Truth and Set. In *The Kleene Symposion*, ed. Jon Barwise et al. North Holland.

Aczel, Peter, and Rachel Lunnon. 1991. Universes and Parameters. In *Situation Theory and Its Applications II*, ed. Jon Barwise et al. CSLI Lecture Notes. Stanford: CSLI Publications.

Barendregt, Henk. 1984. *The Lambda Calculus. Its Syntax and Semantics.* Amsterdam: North Holland.

Barwise, Jon, and Robin Cooper. 1991. Simple Situation Theory and its Graphical Representation. In *Partial and Dynamic Semantics 3*, ed. Jerry Seligman. DYANA-2 Deliverable, No. R2.1.C. Edinburgh: Centre for Cognitive Science.

Bealer, George 1989. On the Identification of Properties and Propositional Functions. *Linguistics and Philosophy* 12.

Chierchia, Gennaro. 1992 Anaphora and Dynamic Binding. *Linguistics and Philosophy* 15.

Chierchia, Gennaro. 1994. Intensionality and Context Change. *Journal of Logic, Language and Information* 3.

Cooper, Robin. 1993. Towards a General Semantic Framework. In *Integrating Semantic Theories.* DYANA-2 Deliverable, No. R2.1.A. Edinburgh: Centre for Cognitive Science.

Dekker, Paul. 1993. *Transsentential Meditations.* Illc dissertation, University of Amsterdam.

Groenendijk, Jeroen, and Martin Stokhof. 1990. Dynamic Montague Grammar. Technical report. Amsterdam: ITLI.

Groenendijk, Jeroen, and Martin Stokhof. 1991. Dynamic Predicate Logic. *Linguistics and Philosophy* 14.

Kamp, Hans. 1981. A Theory of Truth and Semantic Representation. In *Formal Methods in the Study of Language*, ed. Gronendijk et al. Amsterdam: Mathematical Centre.

Turner, Raymond. 1990. *Truth and Modality for Knowledge Representation.* London: Pitman.

Zeevat, Henk. 1991a. *Aspects of Discourse Semantics and Unification Grammar.* Doctoral dissertation, University of Amsterdam.

Zeevat, Henk. 1991b. Static Semantics. In *Partial and Dynamic Semantics 1*, ed. Johan van Benthem. DYANA Deliverable, No. R2.1.A. Edinburgh: Centre for Cognitive Science.

34

Minimal Truth Predicates and Situation Theory

STEPHEN SCHULZ

Introduction

The project I'm sketching here began as an attempt to employ modal situation theory in explicating and investigating anti-realist construals of truth. I focused on Crispin Wright's notion of superassertibility, which is a modification of the Peirce/Putnam notion of that which would be confirmed in the ideal limit of inquiry. This brought up the question of whether situation theory itself was sufficiently realism-neutral to serve as a framework for such an investigation. The second phase of the project, then, centered on a consideration of how one might model the supports relation, situation theory's truth surrogate, anti-realistically. It does appear that one can consistently contravene the traditional assumption that the supports relation is bivalent, and this should be enough to quell any worries about there being a built-in realist bias in situation theory. But there is the further question of whether one *should* construe the supports relation as being epistemically constrained and thus non-bivalent. I want to suggest that this depends on whether one is thinking of the relata in question as possible pieces of information, which deserve the name 'infons', or as more like parts of Stalnakerian possible worlds, which are more happily called 'states of affairs'. A willingness to slide between the two notions would manifest a realist presupposition which appears open to challenge. What sense could it make, after all, to speak of possible pieces of information which are in principle undetectable?

I would like to thank the Moraga conferees, and especially Jerry Seligman, for challenging and helpful discussions.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds

34.1 Superassertibility

34.1.1 Wright's Definition

The Peirce/Putnam line of identifying truth with what would be confirmed in the ideal limit of inquiry may well strike one as dubious. Why should there be a unique ideal limit unless it simply consisted in the godlike possession of *all* the facts, in the realist's sense? How does one make sense of the infinite heirarchy of ideally warranted beliefs about the epistemic status of lower order beliefs or the matter of how a cognizer could have warrant for a belief to the effect that 'this is all the facts there are'? How in general can a pragmatist or any anti–realist make sense of the infinite extension of cognitive capacity which the proposed ideal limit would require? Worries about a unique ideal limit, at least, need not force us to give up on epistemically constrained truth predicate candidates altogether. Wright suggests the notion of superassertibility, which applies to a statement just in case a subject could be in an informational circumstance relevantly similar to the actual world in which the statement would be justifiably asserted and such that no informational improvement would defeat that justification. There is no assumption of an ideal limit, nor would the subject in question necessarily be in a position to *know* that their warrant can not in fact be defeated. Superassertibility is supposed to be a generalization of the sort of warrant provided by (intuitionistically acceptable) mathematical proof. Intuitionistic mathematical practice will be used here as both an urexample and a point of comparison. Quoting the original definition[1]

> 'P' is *superassertible* just in case the world will, in sufficiently favourable circumstances, permit the generation in an investigating subject, S, of a set of beliefs, $\{B_1, ..., B_n\}$ with the following characteristics:
> (a) S has adequate grounds for regarding each of $\{B_1, ..., B_n\}$ as an item of knowledge.
> (b) The status of each of $\{B_1, ..., B_n\}$ as an item of S's knowledge will survive arbitrarily close and extensive investigation.
> (c) The state of information constituted by $\{B_1, ..., B_n\}$ warrants the assertion of 'P'.
> (d) The case provided by $\{B_1, ..., B_n\}$ for 'P' is not, *in fact*, defeasible; i.e. no $\{B_1, ..., B_n, ..., B_z\}$ containing $\{B_1, ..., B_n\}$ and satisfying (a) and (b) for some S, yet *failing* to warrant 'P', can be achieved in this world, no matter how favourable the circumstances for the attempt.

The use of modality in this definition is crucial. What we are looking for ultimately are counterparts of the subject S who warrantedly believe that

[1] See pages 272–279 of Wright 1993.

P in worlds in which the facts relevant to an assessment of P are shielded as much as possible from whatever further facts may be necessary for the existence of such an S counterpart. The degree to which such shielding is possible will depend on the degree to which P is not projected, response–dependent or Euthyphroic. Call a world 'P–independent' (for a subject S) if this sort of shielding is complete so that the case for or against P is informationally isolated from whatever is required for the existence of S. We should not assume that we will have excluded middle for statements of the form 'The actual world is P–independent', for all P.

34.1.2 Warrant

In general, justification is a situation–relative phenomenon, as well as being relative to a subject's background beliefs. I want to distinguish two strengths of warrant: one will be roughly conceptual entailment or what is embodied in intuitionistically acceptable proof moves; the other will be the more usual sort of warrant. Strong warrant is essentially non–holistic and not relative to context, while the second kind is usually both. To define superassertibility we only need garden–variety warrant. Here the idea will be that information will warrant a statement in some situation if the following belief would be justified: that that information's characterizing the situation implies that the information expressed by the statement also characterizes the situation. The notion of justification for individual statements is either taken as primitive or defined in terms of a notion of assertibility. With the definition of warrant I have in mind a typical scientific theorizer who, without thinking that there is anything akin to logical (or analytic) necessity about the connection, comes to believe that an underlying causal mechanism supports the correlation between (observable) information σ and the truth or acceptability of (theoretical) statement P in situation s. Such a theorizer would (typically) employ the standard material conditional[2] to express this connection. Let's try the following formalization:

(W) Let ρ be the information in P. The infon σ *warrants the assertion of P in situation* s iff for any normal subject S, $s \models \ll \mathrm{Bel}\,(\mathrm{S}, \ll s \models \sigma \supset s \models \rho; 1 \gg) \mathrel{\Box\!\!\rightarrow} \mathrm{JBel}\,(\mathrm{S}, \ll s \models \sigma \supset s \models \rho; 1 \gg); 1 \gg$.

There are several ties to the situation s implicit in this. For one thing, the belief in question is not the free–floating belief that there's some sort of

[2] I'm allowing myself the expedient here of appealing to what is apparently a truth–theoretic notion. A similar situation presents itself when I employ the notion of consistency a few paragraphs hence. It would defeat our purpose if the notion of truth were ineliminably appealed to in this exposition. Fortunately, it's not. One may, for example, provide a probabilistic semantics for classical and intuitionistic logic which makes no mention of truth. See Field July 1977 and van Fraassen 1981.

abstract connection between the information and the statement but rather a belief that if this particular situation supports the information then the statement will be true (or acceptable or whatever) in it. Also the presence of s in front of the $\models$ has the effect of privileging the infons which characterize s in the sense that their importance in determining the closeness of worlds relevant to evaluating the counterfactual is paramount. This is a general point about modal situation theory. The idea is that, while there is no clear sense to be made of accessibility and closeness between possible situations, we can make sense of evaluating counterfactuals at a pair consisting of a situation and world it is part of. Closeness is still a matter to be evaluated at the level of worlds, but the situation parameter can be used to weight the importance of the issues decided in that situation, so that the closeness measure between worlds will have as highest priority to make sure that nearby worlds agree with the one we are evaluating from with respect to that situation.

As for representing the idea that justification is situated, this is accomplished simply by exploiting the basic relation between situations and infons. We can express the idea that a subject would, in a certain situation, be justified in believing a certain infon ($s \models \mathrm{Bel}(\mathrm{S}, \sigma) \,\Box\!\!\rightarrow \mathrm{JBel}(\mathrm{S}, \sigma)$) or that a subject in a situation would be justified in believing that a certain infon characterized a certain situation, often but not necessarily the situation they are in ($s \models \mathrm{Bel}(\mathrm{S}, \ll s' \models \sigma;\ 1 \gg) \,\Box\!\!\rightarrow \mathrm{JBel}(\mathrm{S}, \ll s' \models \sigma;\ 1 \gg)$). In formalizing the stability of warrant constraint we need to express that in any situation extending the original one (the one in which the warranting set B is justified) any set of infons extending B which is justified in the extended situation also warrants P.

Finally, we will need a condition which assures us that the set of beliefs which provide the warrant in question is internally consistent. Call a set of beliefs 'epistemically consistent' if its union with other justified beliefs which would be generated by an arbitrarily extensive internal review of the epistemic pedigree of the original set (a review, that is, not involving further gathering of empirical data) would be consistent in one's favorite sense. Since we are talking about sets of infons, consistency would presumably amount to containment in a maximal ideal in the lattice of infons meeting the class of constraints which determine possibility.

With regard to the stability of warrant condition, we must keep in mind that all we've required about the infons in B is that they are believed justifiably and are epistemically consistent: they may well not all be actually true, (or acceptable or superassertible or whatever). In particular, when we extend B with new justifiably believed infons, we may arrive at inconsistent sets. It is not consistency we wish to preserve but rather the strength of the original case made for P by B. If there are new infons B' which would tend to undermine the case for P, we need for it to be that B wins out, in other words, that in the context of $B \cup B'$, B discredits B' rather than vice versa. Even if B is an airtight case for P, there will always be the possibility that some infon may come along which is justified *independently* of B and which tends to undermine P. The airtightness the warrant B gives to P will be manifested not in the impossibility of such infons but rather in their failure to remain justified when put into 'the big picture', that is, when the justifiability of belief in them is considered in a context which forces one to choose between P–undermining B's and the original B.

34.1.3 Superassertibility Refigured

Putting all of this together, we can formulate the following.

Definition. P is *superassertible in situation s* just in case

(S0) There is a nearby (possibly the actual) world w which comes as close as possible to making essentially the same case for P as the actual world, given the existence of subjects and beliefs as required below.

(S1) That world w contains a cognizer S such that JBel (S, σ_i), for each σ_i in $B=\{\sigma_1, \ldots, \sigma_n\}$.

(S2) The set of infons B is epistemically consistent.

(S3) Let σ_B be the conjunction of infons in B and ρ be the information in P. Then $s \models \ll \text{Bel}(\text{S}, \ll s \models \sigma_B \supset s \models \rho; 1 \gg) \mathrel{\Box\!\!\rightarrow} \text{JBel}(\text{S}, \ll s \models \sigma_B \supset s \models \rho; 1 \gg); 1 \gg$.

(S4) For every situation $s' \trianglerighteq s$ and infon $\sigma' \sqsupseteq \sigma_B$, such that $s' \models \text{Bel}(\text{S}, \sigma') \mathrel{\Box\!\!\rightarrow} \text{JBel}(\text{S}, \sigma')$ and σ' is epistemically consistent, we have $s' \models \ll \text{Bel}(\text{S}, \ll s' \models \sigma' \supset s' \models \rho; 1 \gg) \mathrel{\Box\!\!\rightarrow} \text{JBel}(\text{S}, \ll s' \models \sigma' \supset s' \models \rho; 1 \gg); 1 \gg$.

34.2 Idempotence and Negation Equivalence

34.2.1 Challenges to Superassertibility

One may identify a realism–neutral minimal core of intuitions which all parties to the debate(s) over realism can share. One part of this core is the Disquotation Schema. The deflationist typically interprets its *a priori* status as telling us that the two sides of the biconditional have the same cognitive content, while the 'inflationist' will deny this. Another neutral principle is that the class of statements which are appropriately assertoric so as to be minimally truth–apt is closed under certain basic logical operations, in particular, negation and conjunction. Thus, from the DS, we have

(1) 'It is not the case that S' is true iff it is not the case that S.

But negating both sides of ' "S" is true iff S' gives us

(2) It is not the case that ['S' is true] iff it is not the case that [S].[3]

from which we can derive by transitivity the principle of Negation Equivalence

(NE) 'It is not the case that S' is true iff it is not the case that ['S' is true].

Note that there is nothing intuitionistically unacceptable about this derivation. Thus, even though it looks rather like a straightforward enforcement of bivalence, it had better not be, else the Dummett and Wright sort of intuitionistically inclined anti–realist is out of the truth business altogether.

In addition to requiring Negation Equivalence, the Disquotation Schema also demands that a truth predicate be idempotent, that is, for any statement P we have

(I) It is true that it is true that P iff it is true that P.

simply by rewriting ' "it is true that P" is true' as 'it is true that it is true that P'. I hope no one will worry about the expedient shift between truth as a predicate and truth as an operator. Let's begin by considering whether superassertibility is idempotent, whether S(S(P)) iff S(P) for every P, by looking at the left to right implication.

34.2.2 Idempotence $\rightarrow$

Might it be that S(S(P)) without it being the case that S(P)? Roughly, if it is possible to have a stable warrant for its being possible to have a stable warrant for P, does this guarantee that it is possible to have a stable warrant for P? Phrased this way, it looks like a version of the axiom characterizing $S4$ modal logic, which is manifested by the transitivity of the accessibility relation. Here the accessibility relation is what's invoked in clause S0 of the definition above: an accessible world is one which is as P–independent as possible given its containment of an S counterpart[4] with beliefs warranting P, one whose case (relative to S) is as close as possible to the actual case for P given the existence of such an S.

I would prefer to avoid the purely technical response of enforcing the $S4$ axiom by decree. For one thing, one can generate dreadful Sorites type problems exploiting the axiom by adding up a tremendous number of incremental changes. The sort of problem I have in mind here, though,

[3] The [ ] notation is only being employed to clarify scope.

[4] There are a lot of *S*s floating around by now. This one stands for the cognizer in question.

may be illustrated by a different sort of fanciful example. Consider the case of Brian, a brain in a vat who is appeared to crystal–ballishly. Brian has come to regard the crystal ball as a reliable oracle as it has for a long time unerringly predicted events. Perhaps Brian has even formed justified beliefs about what appear to be underlying causal mechanisms which would explain its prescience (whatever the word 'cause' means as uttered by Brian). One day, however, the ball pronounces that Brian is a brain in a vat. Every other of Brian's usual information gathering channels continues to provide evidence suggesting that he is a graduate student in a world graced with extra–mental objects. Nonetheless, the effect of the oracle is such that Brian no longer has an epistemically consistent belief set warranting the statement P he would express by saying 'I am not a brain in a vat'.[5] For such a Brian, it would not be the case that S(P).

Couldn't Brian also reason, however, that there is a possible world making essentially the same case with respect to his envatment but in which his counterpart takes the oracle's one crazy pronouncement as demonstrating its unreliability and even incoherence. Such a Brian counterpart would then have a stable epistemically consistent warranting set for "I am not a brain in a vat". Perhaps this second sort of Brian world is accessible from some world accessible from the 'actual' Brian world (maybe from a world in which Brian has read the first chapter of Putnam's *Reason Truth and History*) but is not accessible from the actual Brian world. So we might end up with S(S(P)) without having S(P).

The line of response I want to advocate is that it is a mistake to think that the second sort of world is in some sense *close to* "making essentially the same case for P as the actual (Brian) world". That thought would rest on a false picture wherein the world, so to speak, does all the work of making a case and the subject simply reads it off, wherein there is a clear divide between perception and cognition, a sort of Myth of the Given. Rather, in our present example, the sceptical Brian worlds should fall well outside the second $\Diamond$, since their case for P is quite non–trivially different from the case made in the credulous Brian worlds. This example has perhaps the perverse and amusing side–effect of exaggerating the epistemic import of *a priori* philosophical reflections.

34.2.3 Relativism

This sort of response generalizes. In doing so, we run across a problem which was already raised by my proposed definition of superassertibility but which I sloughed over. The form of my definition was 'P is superassertible

[5]Note that *this* P is in fact true, as long as we are considering Brian's correlate of what we would express using a homophonic public language utterance. Brian's word 'vat' doesn't, after all, mean vat.

iff there is a possible cognizer who...' From then on everything was relative to a particular cognizer and her counterparts. Now we're emphasizing the role played by these hypothetical cognizers and this brings up the question of the extent to which P's being superassertible should depend on which cognizer we have in mind, what conceptual scheme they employ, which evidential system they subscribe to, what time it is, and so on. Can we be assured that we won't, for some P, have two possible cognizers, one making P superassertible and the other making ~P[6] superassertible? If we can't be sure, then presumably we would be driven to turning superassertibility into a relation along such lines as: 'P is superassertible–for–A–at–*t* iff there is a possible finite extension S of A such that...' Many would feel that by embracing this sort of serious relativism we would have given up the game of arguing that superassertibility can function as a *truth* predicate.

It appears that the problem we're facing here goes beyond what has typically been addressed in situation theory by remarking that everything in the theory presupposes that one has a scheme of individuation at hand. It is primarily the relativistic consequences of possible evidential differences which are of concern here. Might we not, for example, have cognizers employing the same individuation scheme but for whom mutually incompatible evidence is available due to temporal difference?[7] Alternatively, mightn't we have two cognizers with incommensurable conceptual schemes so that we cannot make sense of both of them in terms of one scheme of individuation?[8]

My roughshod way with this matter, which is not intended to denigrate the subtlety of the issues involved, is to split the problem into two cases. The upshot will be that there doesn't, in the end, turn out to be anything more here than what is addressed by the aforementioned usual remark about individuation schemes. Problems of the first sort, exemplified by the evidential import of temporal differences, are to be handled by interpreting the idealization implicit in the definition in a quite liberal way. The spatiotemporal location of the utterer A in question is no more essential to them than their relativistic reference frame: we simply allow

[6]I'm assuming here that we can recognize and make sense of negations, which say essentially that *it is not the case that P*, pre–theoretically and, in particular, independently of any considerations about the nature of truth.

[7]This sort of problem is addressed by Dummett in The Reality of the Past, in Dummett 1978 and by Wright in a number of essays in Wright 1993.

[8]The contrast I have in mind here between a conceptual scheme and a scheme of individuation is that the latter is a vocabulary which, though perhaps largely stipulative, is essentially part of public language, while the former may have only to do with dispositions and mental representation within a single mind. The distinction, and even the sense of talking this way in the first place, has been attacked by Davidson, famously in Davidson 1973–4 and less famously in Davidson 1989.

for counterparts of them in a variety of spatiotemporal locations.[9] There should be nothing perspectival, in this sense, which would result in our having cognizers whose evidential potentials would entail the superassertibility of both P and ∼P. Problems of the second type, in lieu of arguing that the worry is not coherent in the first place, cannot be handled in the same way. One must accept as part of the challenge that we cannot speak of two counterparts of one and the same cognizer employing different conceptual schemes. Conceptual scheme, whatever it is, is essential to a cognizer in a way that spatiotemporal location is not. Here I would insist that the remedy must lie in maintaining the link between content and evidential force in the output just as our rejection of the Myth of the Given insisted on it in the input. If the second sort of worry is coherent, then it should entail that the cognizers in question are not affirming and denying one and the same proposition. Meaning and assertibility conditions can't diverge that far.

34.2.4 Idempotence ←

Does S(P) imply that S(S(P))? If there is a possible stable warranting set for P, must there be a possible stable warranting set for the claim that there is a possible stable warranting set for P? The motivating example may be illustrative here. In formalizing the notion of provability Pr, the property which allows one to derive the second Gödel incompleteness theorem is that

$$\vdash_{\mathrm{PA}} \mathrm{Pr}\,(\|\mathrm{A}\|) \supset \mathrm{Pr}\,(\|\mathrm{Pr}\,(\|\mathrm{A}\|)\,\|).^{10}$$

What we have here is the $S4$ axiom, again, in another guise. Namely, if we think of the $\Box$ as meaning "is provable", which imbues necessity, the condition in question is that $\Box p \supset \Box\Box p$. There is a quick intuitive argument for why one might accept this condition in the mathematical case. Proposition P being provable means that it is possible to give a proof of P. Presumably, if one has in hand a proof of P, one can recognize it as such. One way to demonstrate that something is provable is to display a proof of it. Thus, if it is possible to have in hand a proof of P, it is possible to demonstrate that P is provable. That is, if it is possible to prove P, it is possible to prove that P is provable. That argument is airtight, as long as the crucial premise, that it is always possible to recognize a proof for what it is, is true. In the case of mathematical proofs it is quite plausible

[9]This will no doubt be quite messy to work out. If we accept Kripke's doctrine of the essentiality of origins, then we will have to move the entire geneological tree around in spacetime. It is also hard to see how we could both preserve natural laws, so that these worlds can "make essentially the same case for P as the actual world", and realize the sort of idealizations of cognizers which we will need.

[10]The notation $\|\mathrm{A}\|$ refers to the Gödel number of the formula A (any formula in the language of arithmetic).

since the relation which most naturally represents that Y is a proof of X in formal system S is primitive recursive. But evidence does not generally come with its pedigree stamped on its forehead.

Moving back to superassertibility, we should try to compensate for this fact by availing ourselves of the added degree of freedom allowed us: all that needs to be the case is that if there is an available stable warrant then there is also a stable warrant for the claim that *some set of infons or other* is a stable warrant. If there is in fact a stable warranting set $\{B_i\}$ then there should be some set $\{C_j\}$ which warrants that there is a warranting set $\{B_i\}$, but will it warrant also that $\{B_i\}$ provides *stable* warrant and can we find some $\{C_j\}$ of this sort whose warrant will itself be stable? It seems that very general methodological principles should suffice to support a positive answer to the first question. We couldn't make much sense of the idea of warrant if we didn't take warrants to be stable in the absense of specific reasons to regard them otherwise. A negative answer to the second question would posit the strange situation wherein every $\{C_j\}$ which warrants that there is a $\{B_i\}$ set gets undermined on extension but the $\{B_i\}$ is still a P–warrantor in such an extension. It's hard to make sense of that. So it looks as if Idempotence can be made to go through. An interesting point is that no special epistemic circumstances characterizing particular discourses needed be appealed to.

34.2.5 Negation Equivalence $\rightarrow$

If ~P is superassertible, can we be sure that it is not the case that P is (also) superassertible? This is essentially the relativism problem again. If S(~P) then there is some nearby world w_1[11] in which a set B^1 stably warrants ~P for a cognizer S_1 in some situation s_1 in w_1. Suppose there were also a nearby world w_2 and some set B^2 believed by S_2 in situation s_2 in w_2 which stably warrants P. We would have a contradiction provided we are guaranteed that there is some extension s_2' of situation s_2 in which the union of B^1 and B^2 would be justifiably believed. This would follow from what we might call the Shared Justification Condition, namely that any set of infons, all of which are relevant to P, and which are stably justifiable in one world which makes essentially the same case for P as the actual world should be justifiable in some situation in any other world which makes essentially the same case for P as the actual world. This should follow from whatever it is we mean by "making essentially the same case for and against P as...." Failures would amount to serious relativism of the sort which has already been addressed.

[11]Recall that 'nearby' in this context has the special sense appealed to in condition S0 of the definition of superassertibility.

34.2.6 Negation Equivalence ←

The other direction of the biconditional is where most of the action is. From the failure of S(P) can we conclude S(∼P)? We might have P not being superassertible either because there are no nearby warranting sets or because there are some sets but all of them are unstable. In the first case, whatever warranting sets for ∼P might be available would at least be stable, so the following condition would suffice for the implication:

(W) If there are no nearby warranting sets for P then there is at least one nearby warranting set for ∼P.

I'll come back to W in a moment. The second possibility (availability but instability), in the presence of the Shared Justification Condition, amounts precisely to the condition that above every justifiable infon is a justifiable infon warranting ∼P. This is just the order–theoretic property of the justifiable ∼P warrantors being dense in the justifiable infons in the partial order given by $\sqsupseteq$. In the modal situation theory set–up, an infon forces the double negation of a statement just in case the set of infons which force the statement is dense above that infon. If 'worlds' are $\sqsupseteq$-maximal ideals of coherent infons, then what we would have in the present case is that the set of worlds containing stable ∼P warrantors would be dense in the set of (P–)accessible worlds. What this does not rule out is that the actual world is an ideal in which there is an endless sequence of warranting, overturning, re–warranting, re–overturning, and so on. It's not clear what one should say about this possibility. There doesn't seem to be a general argument to the effect that such a case would *have* to turn out to be one in which the proposition P would not, by the realist's lights, correspond to a determinate state of affairs, but that's what I want to suggest is the case.

The first possibility, that superassertibility of P could fail because of a lack of available warranting sets, came down to condition W. Unlike the case of Idempotence, the plausibility of W is clearly discourse–specific. The more inclined we are to view the subject matter as projected or response–dependent, the more plausible W will be. Internalism about ethics, the claim that no ascription of value can be correct unless the instance is one to which subjects are constitutionally capable, in principle, of responding to as having value, would also support W. Without such a principle, one is thrown back onto the well–worn paths of debate between intuitionists and Platonists in the philosophy of mathematics and between empiricists and realists in the philosophy of science. It should be pointed out, as Wright notes,[12] that W will also have to mark a breakdown of excluded middle if superassertibility is to have a chance of serving as a truth predicate. Unless we admit that we cannot guarantee for every P that either there are nearby

[12]On p. 42 of Wright 1992.

warranting sets or there are not, we would end up enforcing bivalence, and that is surely inconsistent with the notion that truth is epistemically constrained.

34.3 Non–Bivalent 'Supports' Relations

Does it make sense to couch the realism debate with respect to the truth predicate in an ideology one of whose basic notions (the $\models$ relation) is standardly conceived of in a substantively realist way? It might be illuminating to see what effect it would have on situation theory, and in particular on its use as a background for the realism debate, to define a supports relation which is not taken to be bivalent. At the very least, if situation theory ideology is to be useful for the purposes to which I wish here to put it, it must be possible to model the supports relation as being a relation which need not be bivalent. The notion of a supports relation is defined formally in Barwise and Etchemendy 1990 in such a way as not to enforce bivalence. Any relation $\models\ \supseteq Sit \times \mathcal{I}$ is a supports relation so long as (a) if $s \models \sigma$ and $\sigma \Rightarrow \tau$ then $s \models \tau$, (b) $s \not\models \mathbf{0}$ and $s \models \mathbf{1}$, (c) $s \models \sigma \wedge \tau$ if $s \models \sigma$ and $s \models \tau$, and (d) $s \models \sigma \vee \tau$ only if $s \models \sigma$ or $s \models \tau$.[13] If we could eliminate the use of the standard supports relation in the definition of superassertibility, for example, we could try *defining* $s \models \sigma$ as a kind of localized superassertibility. Regular superassertibility could then be seen as a globalization of the supports relation ('there is some situation which supports....') The primitive relation would be assertibility or justified belief relative to (in) a situation, and we would localize the superassertibility notion by restricting the possible informational extensions to those which would remain stably assertible with respect to the situation in question. If we didn't tie it down to the situation in some way, of course, we'd end up with exactly the same infons being supported in every situation.

Some norm of assertibility should, it would seem, be presupposed by our deployment of the notion of assertoric content. Since we get the notion for free, as it were, it looks like a good candidate for a primitive. I remarked above how assertibility was situation–relative. This could be further elaborated, as could specific norms of assertibility–in–a–situation, but my purpose here is to sketch how the outlines of the big picture might go. Let's assume, then, that we have a relation $J(\mathrm{S}, s, \sigma)$ between subjects, situations and infons which says that a statement expressing σ would be assertible for subject S in situation s. For notational convenience we can write $J(\mathrm{S}, s, X)$ in case every infon in the set X is assertible in s.

We will also need a syntactic notion of consistency for infons which isn't based on the standard supports relation. Let's assume for the moment that

[13] See page 44 of Barwise and Etchemendy 1990.

we have at hand a syntactically based entailment relation $\sigma \Rightarrow \tau$. The sort of consistency we have in mind for a set of infons will be the impossibility of deriving **0** from that set together with the set of constraints. Constraints, though normally thought of as expressing relations between (types of) infons, should be expressible as infons. Another notion which we defined using the supports relation, and of which we must now redefine a syntactic analog, is that of warrant. The only connection between infons and situations which we are now taking as primitive is whatever is embodied in the justified belief/assertibility relation $J(\mathrm{S},\ s,\ \sigma)$. This seems to be enough to capture at least the outline of what we mean by a body of information warranting a statement or the infon it expresses. To say that the information σ warrants the assertion of a statement expressing τ in a situation s is just to say that the conditional $\sigma \mathrel{-\!\!\triangleright} \tau$ is assertible in s, in other words, that $J(\mathrm{S}, s, \sigma \mathrel{-\!\!\triangleright} \tau)$. What I have in mind here is not really the familiar truth-functional connective, since that wouldn't make much sense in this context. The meaning of the $\mathrel{-\!\!\triangleright}$ connective, defined in Schulz 1993, is given by the forcing condition $\sigma \Vdash \mathrm{A} \mathrel{-\!\!\triangleright} \mathrm{B}$ iff $\sigma \Vdash {\sim}\mathrm{A} \cup \mathrm{B}$, where A and B are statements, $\cup$ is disjunction, and $\sim$ is given by the condition $\sigma \Vdash {\sim}\mathrm{A}$ iff for all $\sigma' \sqsupseteq \sigma$ we have (not $\sigma' \Vdash \mathrm{A}$). In fact, the family of connectives $\sim$, $\mathrel{-\!\!\triangleright}$, $\cup$, and $\wedge$ corresponds to the intuitionistic connectives. Again, the usual truth-functional connectives have no place here since we are not operating with a notion of truth or with informationally complete objects.

Just as in the definition of superassertibility, we will need to talk about counterfactual situations. In Schulz 1993 I developed the notion of pseudo-situations, which are parts of the maximal ideals in the infon lattice which play the role of possible worlds. They are closed under the class of constraints. The actual pseudo-situations turn out to be just the sets $Facts(s)$, the set of all infons supported by s, and the non-actual ones are what intuitively would be the sets of facts supported by non-actual situations if there were any. The notion does not, however, require the standard $\models$ relation in its formulation. What we need in the present setting is an analog of $J(\mathrm{S},\ s,\ \sigma)$ which will say that S is "contained" in a pseudo-situation and that σ is justified. We will also have to make sense of is the locution "x makes essentially the same case for σ as y" where y is a situation and x is a pseudo-situation. The problem is that we no longer have recourse to the notion of $Facts(s)$. Instead we would have something like $Just(\Sigma)$, the set of infons whose expressions are statements which would be assertible according to the pseudo-situation Σ, or simply $Just(s)$ if we are talking about a situation rather than a pseudo-situation.

I take it that justification or assertibility relations support some counterfactuals. Part and parcel with making sense of 'P is assertible in s' is making sense of locutions like 'P would be assertible if s were different in such a way'. We can thus interpret 'pseudo–situation Σ makes essentially the same case for σ as s does' by comparing $Just(\Sigma)$ and $Just(s)$. Also, since pseudo–situations will come complete with information about which statements are, by its lights, assertible by persons which "inhabit" it, we can express this with the relation $J'(\mathrm{S}, \Sigma, \sigma)$. Given these materials, let's try defining a supports relation along the following lines:

Definition. For any situation s and infon σ, s ***supports*** σ, $(s \mid\models \sigma)$,[14] just in case

(S0′) There is a nearby world containing a pseudo–situation Σ, which comes as close as possible to making the same case for σ as s does, given the existence of subjects and beliefs as required below.

(S1′) The pseudo–situation Σ 'contains' a cognizer S such that $J'(\mathrm{S}, \Sigma, \tau_i)$ for each τ_i in B.

(S2′) The set of infons B is epistemically consistent with respect to the 'syntactic' entailment relation $\Rightarrow$.

(S3′) The conjunction σ_B of the infons in B warrants σ according to Σ, that is, $J'(\mathrm{S}, \Sigma, \sigma_B \rightarrowtriangle \sigma)$.

(S4′) No informational extension $\sigma' \sqsupseteq \sigma_B$ of B by infons in $Just(\Sigma)$ will fail to continue to warrant σ.

Does such a relation satisfy the conditions for being a supports relation? Consider (a), that if $s \models \sigma$ and $\sigma \Rightarrow \tau$ then $s \models \tau$. The $\Rightarrow$ here may or may not be what I was calling the syntactic entailment relation. If it means something like "τ is provable from σ" then if σ is stably warrantable in s then τ should be as well. The $\Rightarrow$ in the original Barwise and Etchemendy 1990 is the informational containment relation I've been writing as $\sqsupseteq$. That notion is introduced using (a) as a partial definition and employing an antecedently understood supports relation. Condition (b) was that $s \not\models \mathbf{0}$ and $s \models \mathbf{1}$. Here we have to make sure that we have the right $\mathbf{0}$ and $\mathbf{1}$ in mind, namely an analog of the provably true and the provably false. A semantic notion like "false in every possible world' won't do because such a falsity may not be recognizable as such. There should be no problem with (c), which was that $s \models \sigma \wedge \tau$ if $s \models \sigma$ and $s \models \tau$. The Shared Justification Condition, along with the other anti–relativistic considerations rehearsed above, may have to be brought to bear in order to make sense of conjunction in general, depending on what sort of role we

[14] I'm using alternate notation to indicate that this is not the usual supports relation.

end up having the idealized cognizer parameter play. The situation with (d), which was that $s \models (\sigma \vee \tau)$ only if $s \models \sigma$ or $s \models \tau$, is trickier. If P is not decidable, then we might end up with $s|\models(\mathrm{P} \vee \sim\mathrm{P})$ but neither $s|\models\mathrm{P}$ nor $s|\models\sim\mathrm{P}$. We need to ensure that $\vee$ is intuitionistic here, and employ the notion of decidable–in–a–situation–s.

Setting up the situation theory apparatus with $|\models$ as the supports relation is, of course, entirely feasible, since the relation is usually taken as a primitive. Whether or not one ought to do so is another question. What shape such a theory would take deserves looking into.

References

Barwise, Jon, and John Etchemendy. 1990. Information, Infons and Inference. In *Situation Theory and its Applications, I*, ed. Kuniaki Mukai Robin Cooper and John Perry. 33–78. CSLI Lecture Notes Number 22. Stanford: CSLI Publications.

Davidson, Donald. 1973–4. On the Very Idea of a Conceptual Scheme. *Proceedings of the American Philosophical Association* 47:5–20.

Davidson, Donald. 1989. The Myth of the Subjective. In *Relativism: Interpretation and Confrontation*, ed. Michael Krausz. 159–172. University of Notre Dame Press.

Dummett, Michael. 1978. *Truth and Other Enigmas.* Oxford University Press.

Field, Hartry. July 1977. Logic, Meaning and Conceptual Role. *The Journal of Philosophy* 379–409.

Schulz, Stephen. 1993. Modal Situation Theory. In *Situation Theory and It's Applications, III*, ed. Peter Aczel, David Israel, Yasuhiro Katagiri, and Stanley Peters. CSLI Lecture Notes Number 37. Stanford: CSLI Publications.

van Fraassen, Bas. 1981. Probabilistic Semantics Objectified: I. Postulates and Logics. *The Journal of Philosophical Logic* 371–394.

Wright, Crispin. 1992. *Truth and Objectivity.* Cambrigde: Harvard University Press.

Wright, Crispin. 1993. *Realism, Meaning and Truth.* Oxford: Blackwell. second edition.

35

Reasoning with Diagrams and Geometrical Constraints

ATSUSHI SHIMOJIMA

Introduction

There are cases in which reasoning with diagrams is superior to purely sentential reasoning, and also cases in which diagrams mislead a reasoner in a way sentences don't. This indicates that diagrams are playing a special role in reasoning, and this special role makes diagrams a sometimes useful, sometimes dangerous, medium of reasoning. My goal is to build an information-theoretic model of diagrammatic reasoning that predicts *both* the advantages *and* the disadvantages of the use of diagrams in reasoning.

The main hypothesis of this paper is already anticipated in the following passage from Barwise and Etchemendy (1990):

> Diagrams are physical situations. They must be, since we can see them. As such, they obey their own set of constraints.... By choosing a representational scheme appropriately, so that the constraints on the diagrams have a good match with the constraints on the described situation, the diagram can generate a lot of information that the user never need infer. Rather, the user can simply read off facts from the diagram as needed. This situation is in stark contrast to sentential inference, where even the most trivial consequence needs to be inferred explicitly.

I extract the follwoing three points from this passage, to frame our general hypothesis: (i) there exists a certain set of geometrical constraints that govern the formation of any diagram, (ii) depending upon the types of semantics we choose, we may or may not be able to exploit these geometrical

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds

constraints to facilitate reasoning, and (iii) the exploitation of geometrical constraints is *the* defining feature of diagrammatic reasoning, from which we can derive all the virtues and vices of diagrammatic reasoning[1].

This paper aims to substantiate this general hypothesis. In sections 1 and 2, I will examine two examples of diagrammatic reasoning, and identify two different patterns in which a reasoner exploits geometrical constraints. We will discuss how the two patterns of constraint-exploitation lead to the often observed advantages of diagrammatic reasoning. In section 3, I will compare diagrammatic reasoning with sentential reasoning, and show that the exploitation of geometrical constraints is in fact a characteristic of diagrammatic reasoning. I will close my discussion by proposing a definition of what it is to use symbols as diagrams, and what it is not to use them as diagrams. In this abstract version, however, I will not discuss the negative aspects of diagrammatic reasoning.

35.1 Exploitation of Geometrical Constraints

I claimed that the characteristic of diagrammatic reasoning is the exploitation of geometrical constraints. But exactly how, if ever, does a reasoner exploit geometrical constraints in reasoning? Let us look at a simple example of diagrammatic reasoning.

Suppose we have the following initial information:

(1) No cats drink whisky.

(2) Jerry is a cat.

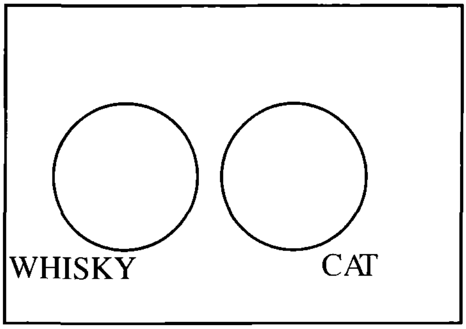
(3) If Jerry does not drink whisky, every cat lives a sober life.

We want to know if Jerry lives a sober life. Using Euler diagrams,[2] we reason in the following way:

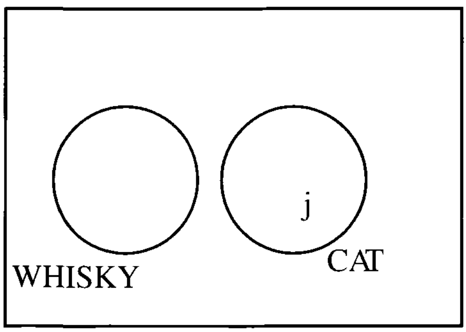
STEP 1. Represent in an Euler diagram the information (1):

[1]To be accurate, Barwise and Etchemendy take the exploitation of geometrical constraints as *one* of many characteristics of diagrammatic reasoning, so that (iii) should be taken as my own strengthening of their original position.

[2]The proof system that I am assuming in this example is the one developed in Hammer and Shin (1995), except that the latter is not intended to express any membership relation between particular objects and sets. Note that, in this system, an overlap between two closed curves does not mean that the corresponding sets have a non-empty intersection.



STEP 2. Encode into the diagram the information (2):



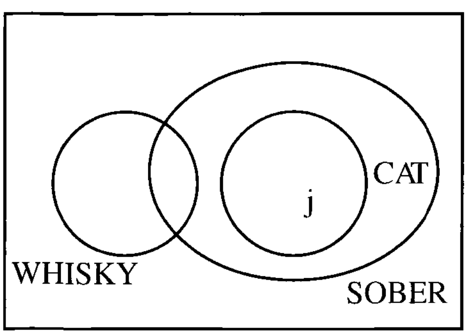
STEP 3. Observe from the diagram that:

(4) Jerry does not drink whisky.

STEP 4. Obtain from (3) and (4) the information:

(5) Every cat lives a sober life.

STEP 5. Encode into the diagram the information (5):



STEP 6. Observe from the diagram that:

(6) Jerry lives a sober life.

Let us focus on steps 1–3 first. In steps 1 and 2, we represent the information (1) and (2) by drawing a diagram in which the following facts hold:

(7) A circle labeled "CAT" and a circle labeled "WHISKY" appear without overlap.

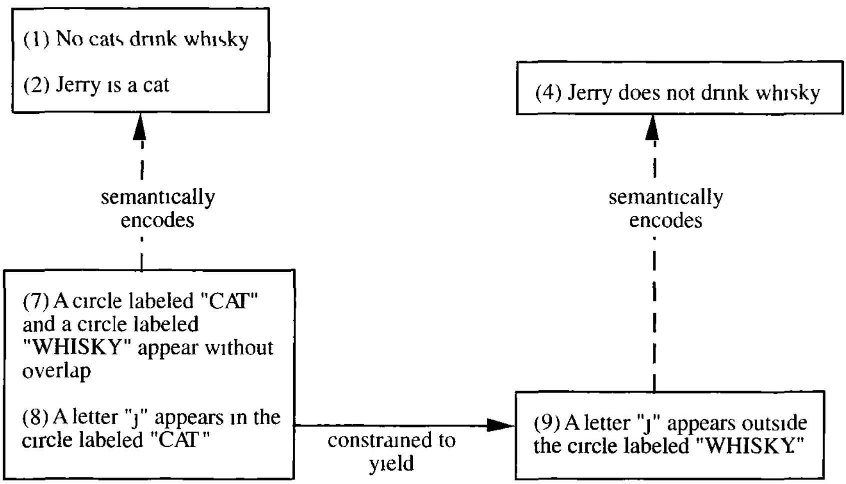
(8) A letter "j" appears in the circle labeled "CAT."

As a result, facts (7) and (8) about the diagram semantically encode facts (1) and (2) about the target situation. Now, whenever we draw two non-overlapping circles and put a letter in one of the circles, the letter must appear outside the other circle. Thus, as a matter of geometrical necessity, facts (7) and (8) yield another fact, (9), about the same diagram:

(9) A letter "j" appears outside the circle labeled "WHISKY."

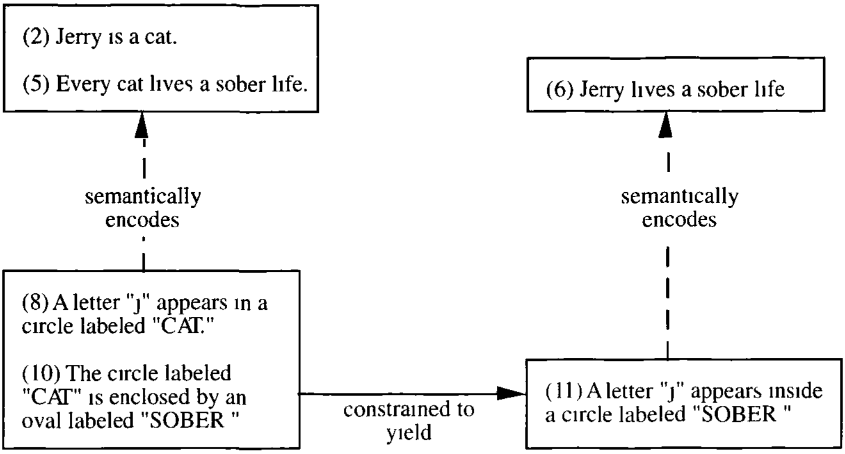
This fact in turn has an independent semantic value, and allows us to read off ("observe") the information (4) from the diagram (step 3).

To display the relationship among the relevant facts:



Note that fact (9) is concurrent with facts (7) and (8), and its production requires no separate manipulation of the diagram. Yet (9) encodes a new piece of information that (7) and (8) do not encode.

Let us look at steps 2, 5, and 6 for a similar example. Here is the relationship among the relevant facts:



To represent facts (2) and (5), we modify our diagram so that fact (10), as well as (8), may hold in it: (steps 2 and 4):

(10) The circle labeled "CAT" is enclosed by an oval labeled "SOBER."

Now, whenever we write a letter in a circle and enclose the circle in an oval, the letter must appear inside the oval. So, as a matter of geometrical necessity, facts (8) and (10) yield another fact:

(11) A letter "j" appears inside a circle labeled "SOBER."

Although this fact is obtained "for free," it has an independent semantic value, and allows us to read off the information (6) from the diagram (step 6).

Generally, to represent a piece of information, we either draw a new diagram or modify an old diagram. If the procedure is successful, a certain fact comes to hold in our diagram that semantically encodes the information to be represented. Now, it is often the case that, by the geometrical constraints, this fact forces an additional fact to hold in the same diagram. Our semantics may or may not assign a semantic value to the latter fact, but if it does, then this additional fact lets us read off new information from the diagram, without costing a separate application of a manual operation. So, if the semantics that we employ lets us exploit geometrical constraints in this way, we can cut down the number of *applications* of structural operation necessary in reasoning.[3] This explains the often observed phenomenon that a derivation in diagrams has fewer steps than the corresponding derivation in sentences.

Theoretically, whenever n applications of structural operations have

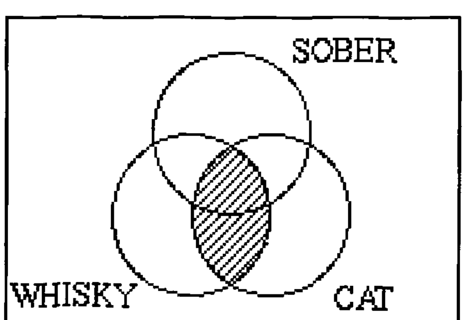
[3]Barwise and Etchemendy have already hinted at this way of exploiting geometrical constraints in the passage cited above.

more than n structural consequences, we can exploit this "increasing" consequence relation to reduce the number of applications of structural operations. From this general point of view, both of our examples are cases in which two applications of structural operations (steps 1 and 2 in the first example; 2 and 5 in the second) have three structural consequences (facts (7), (8), and (9) in the first example; (8), (10), and (11) in the second) with different semantic values. Alternatively, we can see the whole derivation as a case in which three applications of structural operations (steps 1, 2, and 5) have five structural consequences (facts (7)–(11)). Instead of exploring more examples of this pattern, however, let us turn to another, totally different way of exploiting geometrical constraints.

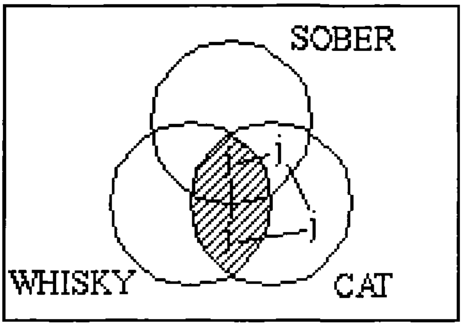
35.2 Exploitation of Geometrical Constraints Continued

In the first sample reasoning, we used Euler diagrams to derive the information (6) from (1)–(3). How do we attain the same goal if we are to use Venn diagrams, rather than Euler diagrams? We would reason in the following way:

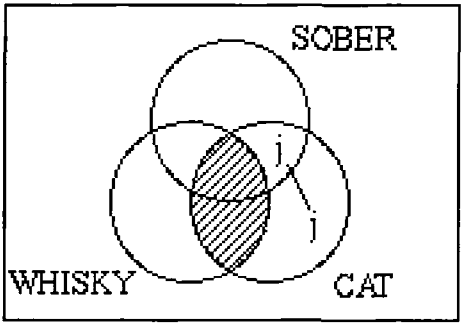
STEP 1. Represent in a Venn diagram the information: (1) no cats drink whisky:



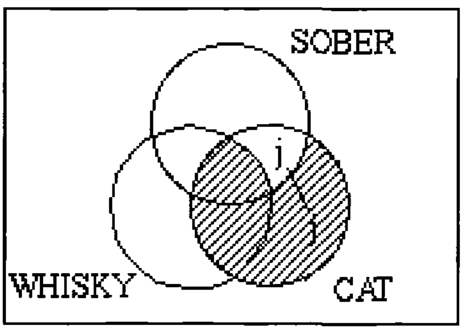
STEP 2. Encode into the diagram the information that: (2) Jerry is a cat:



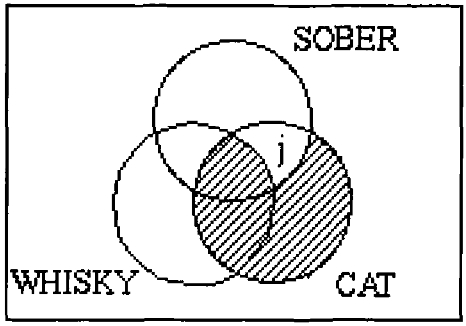
STEP 3. Erase all the shaded letters along with the chains touching them:



STEP 4. Observe from the diagram that: (4) Jerry does not drink whisky.
STEP 5. Obtain from (3) and (4) the information that: (5) every cat lives a sober life.
STEP 6. Encode into the diagram the information (5):



STEP 7. Erase all the shaded letters along with the chains touching them:



STEP 8. Observe from the diagram that: (6) Jerry lives a sober life.

Let us focus on steps 1–4 first. In steps 1 and 2, we represent facts (1) and (2) by drawing a diagram in which:

(12) The intersection of a region labeled "CAT" and a region labeled "WHISKY" is shaded;

(13) A j-sequence appears inside the region labeled "CAT," with a "j" in each minimal sub-region.[4]

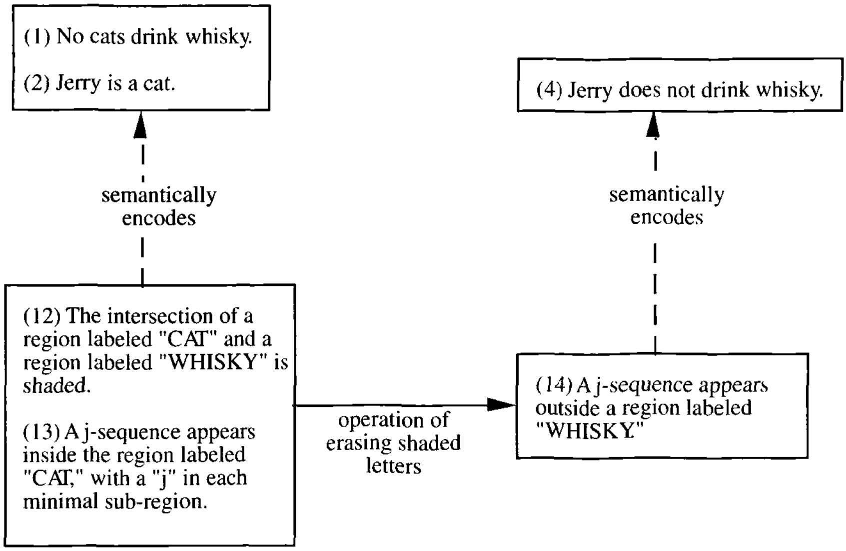
In step 3, we apply a structural operation to the diagram, namely, the operation of erasing shaded letters. Then, as a matter of geometrical necessity, the following fact holds:

(14) A j-sequence appears outside a region labeled "WHISKY,"

which semantically encodes the information that: (4) Jerry does not drink whisky.

To show the relationship among these facts schematically:

[4] A *region* is an area in a diagram that is enclosed by curves and lines. A *minimal region* is a region in which no other region is included. For any letter α, *α-sequence* is a finite sequence of α's connected by lines. I borrow these terms from Shin (1991).



Let us now turn to steps 6–8. We have seen that fact (14) was a geometrical consequence of erasing the shaded letters "j" in step 3. Actually, there is another geometrical consequence of that move, that is:

(15) A j-sequence appears in the complement of the "WHISKY" region relative to the "CAT" region, with a "j" in each minimal sub-region.

In step 6, we represent the information (5) by modifying diagram so that the following fact may hold in it:

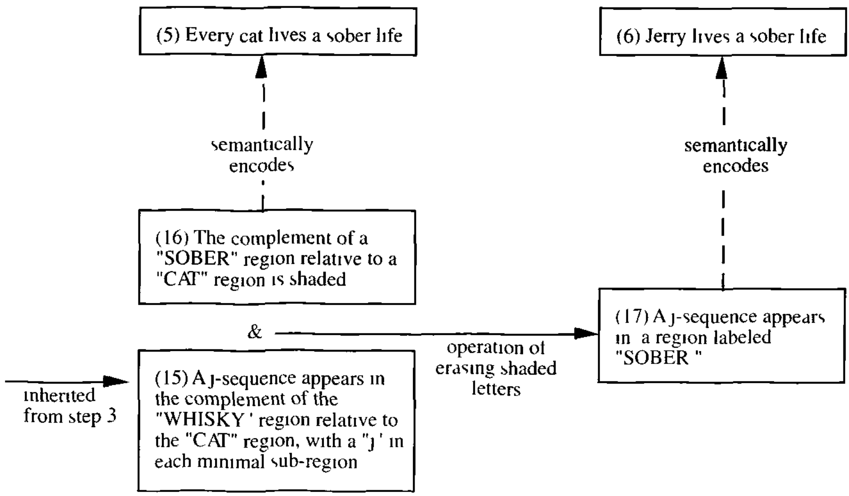
(16) The complement of a "SOBER" region relative to a "CAT" region is shaded.

In step 7, we apply a manual operation of erasing shaded letters again. Then, by a geometrical constraint, the following fact holds in the diagram:

(17) A j-sequence appears in a region labeled "SOBER."

This fact has an independent semantic value on our semantics, and allows us to read off from the diagram that: (6) Jerry lives a sober life.

To display the relationship among the relevant facts:



The point of this example is that, under geometrical constraints, a single structural operation can have different structural consequences, depending on the initial conditions under which it is applied. In our example, the operation of erasing shaded letters has consequence (14) when it is applied under conditions (12) and (13), while the same operation has consequence (17) when applied under conditions (15) and (16). Since our semantics assigns different semantic values, (4) and (6), to these structural consequences, the same operation ends up with different semantic consequences in different occasions. In general, if the semantics that we employ lets us exploit geometrical constraints in this way, we can cut down the number of *types* of operation necessary in reasoning. This explains the phenomenon that a derivation in diagrams often requires a fewer varieties of structural operations than its sentential counterpart. In the above derivation, for example, we just kept erasing shaded letters (aside from representing information into diagrams).

35.3 Comparison to Sentential Reasoning

To summarize what I claimed so far: (i) there is a set of geometrical constraints that govern the formation of any diagram, in the sense that (ii) there are certain *regular*, and therefore *exploitable* connections between applications of structural operations to diagrams and their structural consequences; (iii) in particular, there are cases in which n applications of structural operations have more than n structural consequences; (iv) by taking advantage of this "increasing" consequence relation, we can reduce the number of *applications* of operations necessary in reasoning; (v) there are also cases in which different applications of a single structural operation have different outcomes, depending upon the conditions under which the operation is applied, and (vi) by taking advantage of this one-to-many correspondence from types of operations to the consequences of their applications, we can reduce the number of *types* of operations necessary in reasoning.

A natural question at this point is: "Is the exploitation of geometrical constraints described in (vi) and (v) really a characteristic of diagrammatic reasoning? Don't we do the same things in sentential reasoning?" The goal of this section is to make it clear that, at least in sentential reasoning, the two forms of exploitation of geometrical constraints described in (iv) and (vi), or anything similar to them, are *intentionally* avoided.

The structural operations in sentential reasoning are all productions (writing or printing) of particular sentences. These "sentential operations" are not exceptions to geometrical constraints—there is certain regularity between applications of sentential operations and their structural consequences. Now, writing down a particular sentence is usually constrained to have more than one structural consequence, just as an application of a structural operation to a diagram usually does. In sentential reasoning, however, reasoners are instructed to ignore most of these consequences—the semantics does not assign any semantic values to them. The only structural consequence that has a semantic value is the fact that a particular sentence is now a part of our representation, and its semantic value is simply the semantic content of that sentence.

To wit, suppose that, after writing down the sentence "No cats drink whisky," you write down the sentence "Jerry is cat" just below the first sentence:

No cats drink whisky.
Jerry is a cat.

This structural operation (production of the second sentence) has many structural consequences. First, it has the structural consequence that the sentence "Jerry is cat" is now a part of your representation. Apart from this, the following are all structural consequences of this operation (in this particular application): the word "Jerry" now appears below the word "cats"; the letter "k" now appears above the letter "a"; the two sentences form a shape of inverse trapezoid, etc. But does the operation have more than one semantic effect for this reason? No. These additional consequences are all irrelevant from the semantic standpoint; the only structural consequence *to which our semantics assigns a value* is the first one, and the value is simply the information that Jerry is a cat. It never happens that one of the other structural consequences also has a semantic value, and allows us to read off information other than the one directly encoded in the added sentence.

This is in stark contrast to diagrammatic reasoning, where the semantics is so designed that a single application of a structural operation may have more than one structural consequence with different semantic values. The semantics for sentential reasoning is not designed in that way—it is so designed that, to each *application* of a structural operation, there may correspond at most one structural consequence with a semantic value. In other words, it is designed *not* to exploit geometrical constraints in the way described in (iv).

What about the exploitation of geometrical constraints described in (vi)? It is true that a single sentential operation has different sets of structural consequences in different occasions of application. But it has a fixed structural consequence, regardless of differing conditions of application, when it comes to the structural consequences *with semantic values*.

To wit, suppose you write down the sentence "Jerry is a cat" under two different conditions. In case 1, the sentence "No cats drink whisky" appears just above the area in which you are about to write that sentence, while in case 2, the sentence "Every cat lives a sober life" appears instead:
Case 1

No cats drink whisky.
Jerry is a cat.

Case 2

Every cat lives a sober life.
Jerry is a cat.

Obviously, the same operation (the operation of writing down the sentence "Jerry is a cat") has different sets of structural consequences in these two cases. In case 1, for example, it has the consequence that the word "cat" comes below the word "whisky" while in case 2, it has the consequence that the word "cat" comes below the word "sober." But does the operation have different semantic effects for this reason? No, because neither of these structural consequences has a value on our semantics. In both cases, the only structural consequence that has a semantic value is the bare fact that the sentence "Jerry is a cat" is produced, and hence the information that this operation encodes into our representation is one and the same in the two cases, namely, the information that Jerry is a cat.[5]

[5]One might think that the operation in case 1 encodes into our representation the information that Jerry does not drink whisky, while the same operation in case 2 encodes the information that Jerry lives a sober life. But, strictly speaking, neither piece of information is encoded in our representation until the sentence "Jerry does not drink whisky" or the sentence "Jerry lives a sober life" is written down.

Thus, although a single operation may have different sets of structural consequences under different conditions, the structural consequence *that has a semantic value* does not vary. This is again in stark contrast to diagrammatic reasoning in which the semantics is so designed that different applications of a single structural operation may have different structural consequences with different semantic values. The semantics for sentential reasoning is not designed in that way—it is so designed that, to each *type* of structural operation, there may correspond at most one structural consequence with a semantic value (regardless of differing conditions of application). In other words, it is designed not to exploit geometrical constraints in the way described in (vi).

Altogether, in sentential reasoning there is (vii) a functional correspondence from *applications* of structural operations to structural consequences with semantic values, and (viii) a functional correspondence from *types* of structural operations to structural consequences with semantic values. From one point of view, these are troublesome features that make reasoning more tedious and less spontaneous. From another point of view, however, (vii) and (viii) are advantages. Because of (vii), we do not have to worry that a given application of an operation may have semantic side-effects that we are not aware of. Because of (viii), we do not have to worry that a given type of operation may have different semantic effects in different occasions of application. In sentential reasoning, therefore, it is less likely that, in applying an operation, we happen to encode in our representation a piece of information other than the one that we intend. We have a greater control over our own reasoning, and that is a strength of sentential reasoning, namely, reasoning with features (vii) and (viii).[6]

35.4 Final Remarks

One of the most basic, long-standing questions in the study of diagrammatic reasoning is: "What is a diagram? What distinguishes diagrams from other kinds of symbols, such as sentences?" My account of the characteristic of diagrammatic reasoning suggests an answer. There are no properties intrinsic to symbols that make them diagrams. Accordingly, there is no set of symbols that are diagrams. There are only diagrammatic and non-diagrammatic *uses* of symbols. A person *uses* a set of symbols *as diagrams* if and only if the semantics that he adopts for them allows him to exploit geometrical constraints in one of the two ways described in (iv) and (vi). Again, a symbol is neither a diagram nor a non-diagram itself; only a particular use of it can be diagrammatic or not.

---

[6] I am indebted to Prof. Syun Tutiya for keeping reminding me of the advantage of sentential reasoning.

In this paper, we have studied the characteristic of diagrammatic reasoning mainly in its connection to the observed advantages of diagrammatic reasoning. My overall goal, however, is to let my account also predict the observed disadvantages of diagrammatic reasoning. The fuller version of this paper will attempt it, along with the formalization of the arguments presented here.

References

Barwise, Jon, and John Etchemendy. 1990. Visual Information and Valid Reasoning. In *Visualization in Mathematics.* Washington, DC: Mathematical Asociation of America.

Hammer, Eric, and Sun-Joo Shin. 1995. Euler and the Role of Visualization in Logic. In *this volume.*

Shin, Sun-Joo. 1991. A Situation-Theoretic Account of Valid Reasoning with Venn Diagrams. In *Situation Theory and Its Applications, 2.* Stanford: CSLI Publications.

36

A Legal Reasoning System Based on Situation Theory

SATOSHI TOJO AND STEPHEN WONG

Abstract

Legal reasoning systems research is a new field attracting both AI researchers and legal practitioners. The purpose of this paper is to introduce a formal model of legal reasoning, based on situation theory. On that abstract model, we show an example of reasoning system implemented in a knowledge-base management language $\mathcal{QUIXOTE}$, regarding the language as a situated inference system.

Introduction

Legal reasoning systems research is a new field which has attracted researchers from both the legal and AI domains. Most legal reasoning systems draw arguments by interpreting judicial precedents (old cases) or statutes (legal rules), while more sophisticated systems include both kinds of knowledge. Surveys of the leading projects can be found in Rissland 1991a, Rissland 1991b, Sergot 1990, Gardner 1987, and Ashley 1990.

Thus far, those legal reasoning systems seem to have had weak foundation in formalization, and they have been *ad hoc* combination of various forms of logical inference. Our prime purpose of this paper is to give a sound foundation to legal reasoning system in terms of situation theory in Barwise and Perry 1983, Barwise 1989, and Devlin 1991. And, in addition, we implement this model into a computational form in Knowledge Base Management System (KBMS) $\mathcal{QUIXOTE}$, introduced in Yokota and Yasukawa 1992, Yokota et al. 1993. Regarding the concept of *module* of $\mathcal{QUIXOTE}$ as situation, we show that the language can work as a

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

situated inference system. The set of knowledge bases includes a dictionary of legal ontologies, a database of old cases, and a database of statutes.

The organization of this paper is as follows. Section 1 describes the formulation of legal knowledge at the abstraction level using the theory of situations. Section 2 illustrates the realization of this formulation at the KBMS level using $\mathcal{QUIXOTE}$, and its situated inference mechanisms. The last section concludes this paper.

36.1 Situation Theory for Legal Reasoning

This section introduces a formal model for legal reasoning, especially, penal code, at the abstraction level. The formulation is based on *situation theory*, so we call it a situation-theoretic model ($\mathcal{SM}$), as in Wong 1992.

36.1.1 General Terms

The ontologies of $\mathcal{SM}$ include objects, parameters, relations, infons, and situations. An object designates an individuated part of the real world: a constant or an individual in the sense of classical logic. A parameter refers to an arbitrary object of a given type. An n-placed relation is a property of an n-tuple of argument roles, $r_1, \cdots, r_n$, or slots into which appropriate objects of a certain type can be anchored or substituted. For example, we can define 'eat' as a four-place relation of *Action* type as:

$$< \text{eat}{:}Action \mid eater{:}ANIMAL,\ thing\text{-}eaten{:}EDIBLE\text{-}THING,$$
$$location{:}LOC,\ time{:}TIM >$$

where *eater, thing-eaten, location,* and *time* are roles and the associated types, *ANIMAL* denotes the type of all animals, *EDIBLE-THING* denotes the type of all edible substances, and *LOC* and *TIM* are types of spatial and temporal location.

An infon σ is written as $\ll Rel, a_1, ..., a_n, i \gg$, where *Rel* is a relation, each argument term a_k is a constant object or a parameter, and i is a polarity indicating 1 or 0 (true or false). If an infon contains an n-place relation and m argument terms such that $m < n$, we say that the infon is *unsaturated*; if m = n, it is *saturated*. Any object assigned to fill an argument role of the relation of that infon must be of the appropriate type or must be a parameter that can only anchor to objects of that type. An infon that has no free parameters is called a *parameter-free* infon; otherwise, it is a *parametric* infon. If σ is an infon and f is an *anchor* for some or all of the parameters that occur freely in σ, we denote, by $\sigma[f]$, the infon that results by replacing each v in the domain of f that occurs freely in σ by its value (object constant) $f(v)$. If I is a set of parametric infons and f is an anchor for some or all of the parameters that occur freely in I, then $I[f] = \{\sigma[f] \mid \sigma \in I\}$.

$\mathcal{SM}$ is a triplet $\langle \mathcal{P}, \mathcal{C}, \models \rangle$, where $\mathcal{P}$ is a collection of abstract situations including judicial precedents, a new case, c_n, and a world, w, that is a unique maximal situation of which every other situation is a part. $\mathcal{C}$ is a concept lattice in which objects are introduced and combined with the *subsumption* relation ('=<'), that is an *is-a* relation intuitively, each other. 'An object of a type' is interpreted as 'an object is subsumed by another object corresponds to that type'. We give an example of a concept lattice in Fig. 1. '$\models$' is the support relation, and our interpretation is:

Definition 1 (Supports Relation) *For any $s \in \mathcal{P}$, and any atomic infon σ, $s \models \sigma$ if and only if (iff) $\sigma \in s$.* □

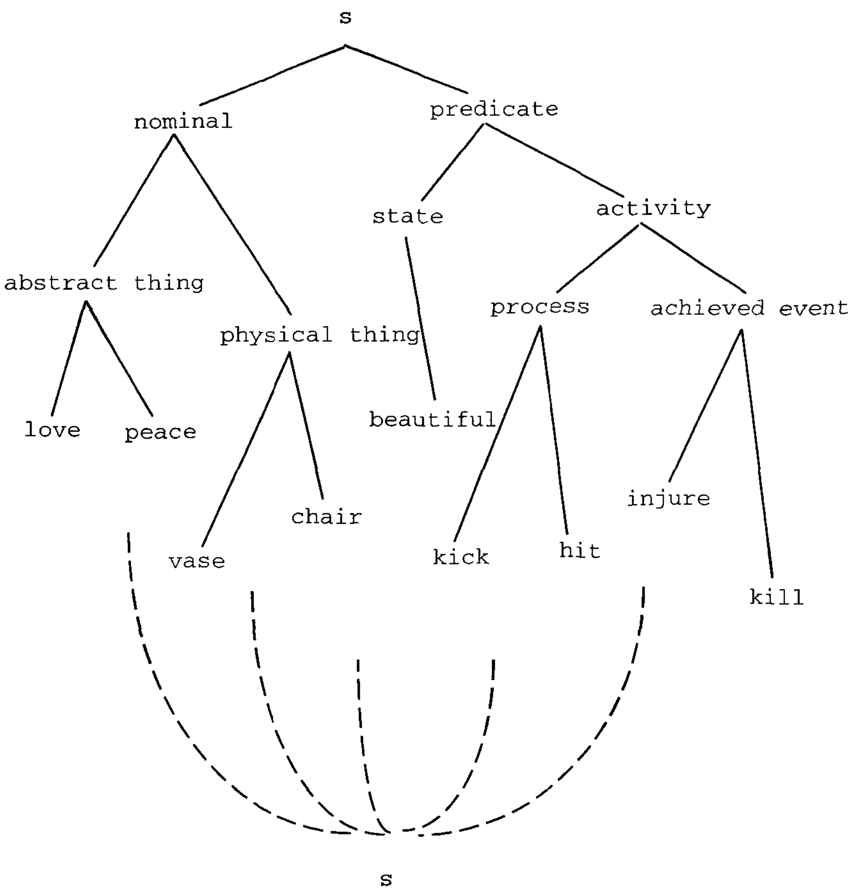


FIGURE 1 Concept lattice

36.1.2 Situated Inference Rules

Reasoning in law is a rule-based decision-making endeavor. A legal reasoning process can be modeled as an inference tree of four layers. The bottom layer consists of a set of basic facts and hypotheses, the second layer involves case rules of individual precedents, the third layer involves case rules which are induced from several precedents or which are generated from certain legal theories, and the top layer concerns legal rules from statutes. An individual or local case rule is used by an agent in an old case to derive plausible legal concepts and propositions. These rules vary from case to case, and their interpretation depends on the particular views and situational perspectives of the agents. An induced case rule has a broader scope and is generalized from a set of precedents. Legal rules are general provisions and definitions of crimes. They are supposed to be universally valid in the country where they are imposed, and neutral. That is, the applicability of these rules is independent from the view of either side (plaintiff or defendant) and every item of information (infon) included is of equal relevance. Though it rarely happens, it may be possible for an agent to skip one or two case rule layers in attaining a legal goal. In Fig. 2, we show the chaining of these various level of rules.

In such a rule-oriented legal domain, *situated inference* has the following general form:

Rule 1 (General Rule) $s_0 \models \sigma_0 \Leftarrow s_1 \models \sigma_1, s_2 \models \sigma_2, ..., s_n \models \sigma_n/B$, *where* $\sigma_0, \sigma_1, ..., \sigma_n$ *are infons, and* $s_o, s_1, ..., s_n$ *are situations.* □

This rule can be read as: "if s_1 supports σ_1, s_2 supports σ_2, and so on, then we can infer that s_0 supports σ_0 under the background conditions or constraints B." $s_0 \models \sigma_0$ is called the head of the rule while the remainder is called the body of the rule. The background conditions, B, are required to be coherent and satisfied before execution of the rule. Note that $c \models I/B$ implies that $c \cup B \models I$, where $c \models I$ as a shorthand for $c \models \sigma_1, c \models \sigma_2, ..., c \models \sigma_n$.

We are particularly interested in three rule instances: local case rules, induced case rules, and legal rules. A local rule is as follows:

Rule 2 (Local Rule) *For* $c \in \mathcal{P}$, cr: $c \models \sigma \Leftarrow c \models I/B_{cr}$. □

where I is called the antecedent of the rule, σ is the consequent, and cr is the label of the rule, which is not part of the rule but which serves to identify the rule. Sometimes, we simply write cr: $c \models \sigma \Leftarrow I/B_{cr}$. Both σ and I are parameter-free. One unique feature of rules in the legal domain is that the consequent is not disjunctive and often a single predicate. The reliability and the scope of application of a local rule will be subject to a set of *background conditions*, B_{cr}. The conditions include information such

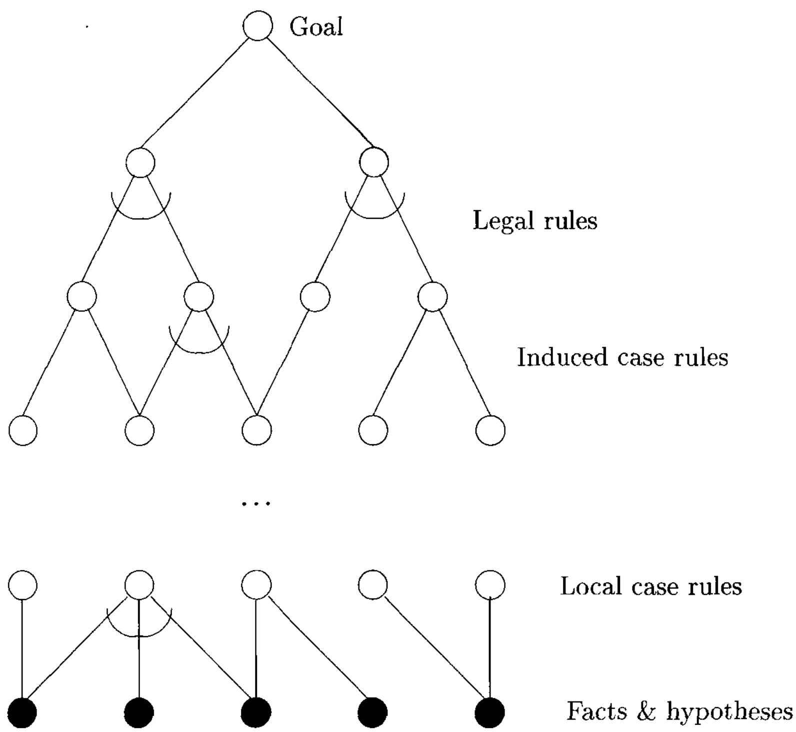


FIGURE 2 A legal inference tree

as an agent's goal and hypotheses; these are crucial in debate to establish the degree of certainty and the scope of applicability of that rule. Usually, it becomes necessary to take background conditions into account and investigate what they are. Many case rules exist in one case and often yield incompatible conclusions. But, the background conditions clarify their hypotheses and perspective. When there is no danger of conclusion, we can write such a rule without stating its background conditions.

Another form of case rule is generalized or induced from several precedents. Owing to its generic nature, an induced case rule is represented as a constraint between two parametric infons, rather than parameter-free ones. Denote I' and σ' as a set of parametric infons and a parametric infon, respectively, such that all parameters that occur in the latter also appear in the former. An induced rule is written as:

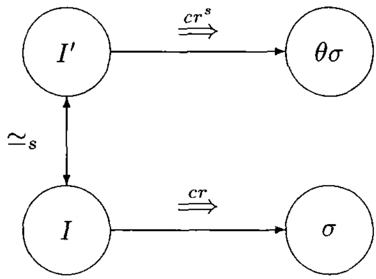


FIGURE 3 Case substitution

Rule 3 (Induced Rule) *For any* $c_1, ..., c_k \in \mathcal{P}$, $c = c_1 \cup c_2, \cup ... \cup c_k$, $ir\colon c \models \sigma' \Leftarrow I'/B_{ir}$. □

where c is coherent and ir is the rule label. Similarly, a legal rule is:

Rule 4 (Legal Rule) $lr\colon w \models \sigma' \Leftarrow I'/B_{lr}$. □

where lr is the rule label and B_{lr} states the background legal theory, such as the aim of punishment or the aim of crime prevention, but not both. Such information is crucial in interpreting the antecedent infons.

36.1.3 Substitution and Anchoring

When a situation of a new case, c_n, supports a similar antecedent of a local rule of c_o, one can draw a conclusion about the new case that is similar to the consequent of that rule.

Definition 2 (Local Rule Substitution) *For* $c_n, c_o \in \mathcal{P}$, $cr^s : c_n \models \sigma\theta$ *if* $cr\colon c_o \models \sigma \Leftarrow I/B_{cr}$ *and* $c_n \models I'/\{B_{cr}\theta \cup B_n\}$ *such that* $I' \simeq_s I$. □

where cr^s is the label of the new rule, B_n is the original background of I' of the new case, and the combined condition after the substitution, $B = B_{cr}\theta \cup B_n$, is coherent. The notation $\simeq_s$ denotes the matching relation between two situations. Section 2 discusses how such a matching is implemented in *QUIXOTE*. The function θ forms a *link* that connects c_n with c_o. This function replaces all terms (objects and relations) in σ and B_{cr} that also occur in I with their matched counterparts in I'. Normally, the background conditions are not included. In Fig. 3, we show the case substitution.

To combine the conclusions supported by different situations, the background conditions of both conclusions must be compatible. That is why

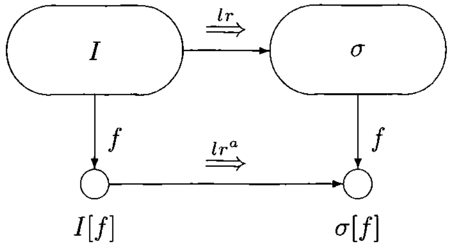


FIGURE 4 Legal anchoring

the background conditions of Rule 1 must be coherent. Rather than substitution, a consequent is derived from a legal rule or an induced rule via anchoring.

Definition 3 (Induced Rule Anchoring) *For* $c_n, c_1, ..., c_k \in \mathcal{P}$ *such that* $c = c_1 \cup c_2 \cup ... \cup c_k$, $ir^a : c_n \models \sigma[f]$ *if* $ir : c \models \sigma \Leftarrow I/B_{ir}$ *and* $c_n \models I[f]/\{B_{ir}[f] \cup B_n\}$. □

Definition 4 (Legal Rule Anchoring) *For* $c_n \in \mathcal{P}$, $lr^a : c_n \models \sigma[f]$ *if* $lr : w \models \sigma \Leftarrow I/B_{lr}$ *and* $c_n \models I[f]/\{B_{lr}[f] \cup B_n\}$. □

In Fig. 4, we show the legal anchoring.

36.1.4 Matching of Infons and Situations

In order to compare the similarity of a new case with precedent cases, we formalize the infon matching and the situation matching. Suppose that a concept lattice is given, where the subsumption relation ('=<') is defined between concepts. $R(\sigma)$ is a function that extracts '*rel*' from an infon σ.

Definition 5 (Infon Matching) *For any two infons* σ_1 *and* σ_2,

1. *If there is a* $R(\sigma_3)$ *such that* $R(\sigma_1)$ =< $R(\sigma_3)$, *and* $R(\sigma_2)$ =< $R(\sigma_3)$ *in a given concept lattice, then* σ_1 *and* σ_2 *are interpreted as* weakly matched infons.
2. *If* $R(\sigma_1) = R(\sigma_2)$, *then then* σ_1 *and* σ_2 *are interpreted as* partially matched.
3. *If all the objects that constitute two infons are identical, then the infons are* exactly matched. □

We give concepts of situation matching below. Note that the concept of situation matching is independent of the concept of infon matching, so that we can apply any definition of infon matching for the following definition of situation matching.

Definition 6 (Situation Matching) *For situations s_1 and s_2,*

1. *If, for every infon in s_1, there is an infon that can match it in s_2, and vice versa, then the two situations are interpreted as* exactly matched *situations.*
2. *For any σ_1 in s_1, there is an infon σ_2 in s_2 that can match σ_1, situation s_1 can be* partially matched *with situation s_2.*
3. *For any σ_1 in s_1 whose* relevance *value is larger than a given threshold level, there is an infon σ_2 in s_2, that can be matched with σ_1, s_1 can be* partially matched *with s_2* w.r.t. relevance value □

Thus, situation matching is a one-way relation in situations, as to whether one situation is *embeddable* into another situation (Fig. 5). Among several matching definitions, we will adopt *weakly matching* for infons and *partially matching w.r.t. relevance value* for situations, in implementation of the following section, for practical reasons. Let us consider the following pair of descriptions:

$$\begin{array}{rll} s_{new} & \models \{ & \ll abandon, mary^{agent} \gg, \\ & & \ll leave, mary^{agent}, june^{object} \gg \; \} \\ s_{old} & \models \{ & \ll abandon, jim^{agent}, tom^{object}, 3^{relevance} \gg, \\ & & \ll leave, jim^{agent}, tom^{object}, 2^{relevance} \gg, \\ & & \ll poor, jim^{agent}, 1^{relevance} \gg \; \} \end{array}$$

If the threshold value is 2, then s_{new} can be *partially matched* with s_{old} *w.r t. the value 2.* On the other hand, if 1 is the threshold instead, s_{new} cannot be *partially matched* with s_{old} *w.r.t. the value 1.* s_{old} cannot be matched with s_{new} in any sense because there is no infon in s_{new} that corresponds to the infon $\ll poor \gg$ in s_{old} (Fig. 6).

36.2 Modeling of Legal Knowledge in *QUIXOTE*

This section introduces the language of *QUIXOTE* and shows how this language can represent the $\mathcal{SM}$ concepts in computable form. A typical *QUIXOTE* database includes the following data structures: (i) the subsumption relations among basic objects, (ii) the submodule relations among modules, and (iii) rules. Our legal reasoning system consists of three databases: a legal dictionary, cases, and statutes. Accordingly, we first introduce the *objects* and *modules* of *QUIXOTE* and explain the data structure of the legal dictionary, then describe the use of *QUIXOTE rules* to represent case-based rules and statutes.

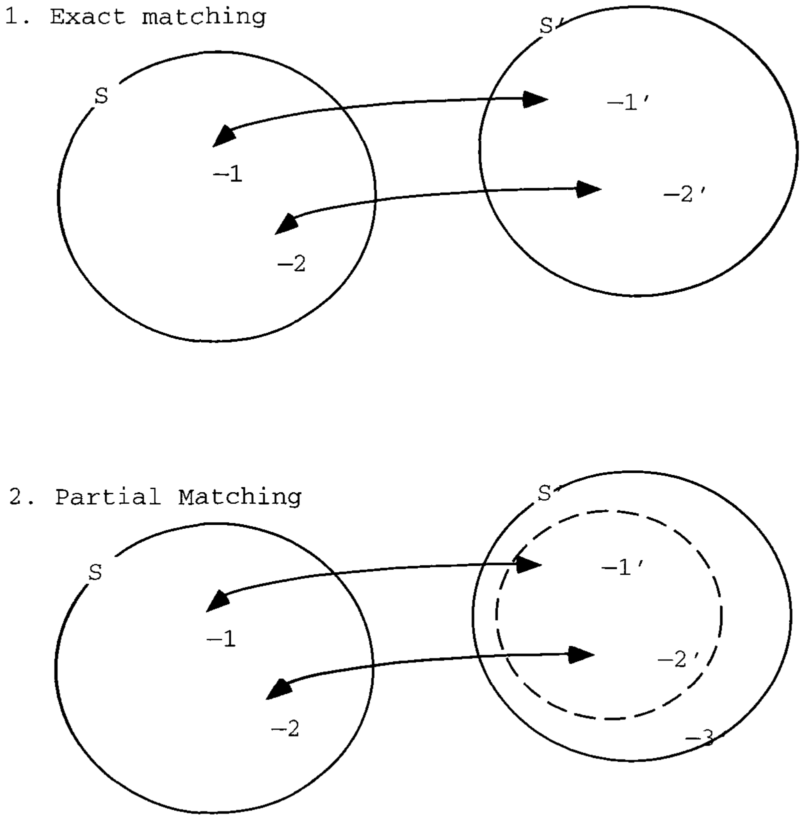


FIGURE 5 Exact and Partial situation matching

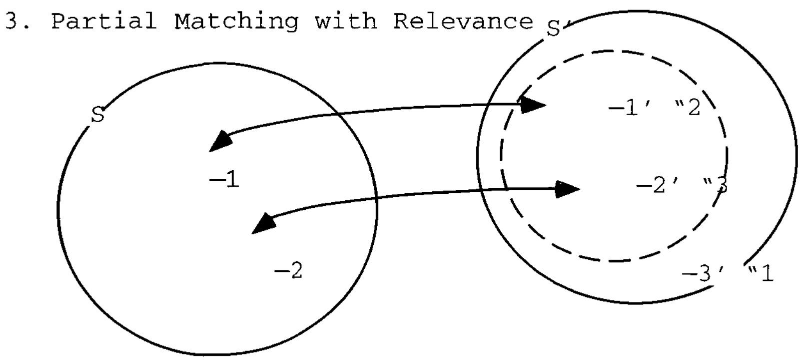


FIGURE 6 Partial situation matching with relevance value

In $\mathcal{QUIXOTE}$, the concepts of $\mathcal{SM}$ are rephrased as follows:

$\mathcal{SM}$	$\mathcal{QUIXOTE}$
situation	module
infon	attribute term
relation name	basic object
type	subsumption
role	label
supporting relation ($\models$)	membership in module (:)

36.2.1 Description of Case and Rule

A $\mathcal{QUIXOTE}$ rule has the following form (compare with Rule 1):

$$\overbrace{m_0 :: H}^{head} \mid \overbrace{HC}^{head\_constraints} \Leftarrow \overbrace{m_1 : B_1, \ldots, m_n : B_n}^{body} \parallel \overbrace{BC}^{body\_constraints} \quad ;;$$

where H or B_i are objects, and HC and BC are sets of formulas (constraints) using subsumption relations. Intuitively, this means that if every B_i holds in a module m_i under the constraints BC, then H and constraints HC hold in m_0. The head constraints and module identifiers can be omitted, and the body constraints, BC, of a rule then constitute the background conditions for that rule.

This study regards a case as being a situation, that is, a set of anchored sentences. Below, we describe a case which is a simplified description of an actual precedent, mentioned in Nitta et al. 1992, Nitta et al. 1993.

Mary's Case
On a cold winter's day, Mary abandoned her son Tom on the street because she was very poor. Tom was just 4 months old. Jim found Tom crying on the street and started to drive Tom by car to the police station. However, Jim caused an accident on the way to the police station. Tom was injured. Jim thought that Tom had died in the accident and left Tom on the street. Tom froze to death.

This aforementioned case contains some human objects and several events with different relevancy. The order of values of the relevance attribute is represented by a subsumption relation, (`l1 =< l2 =< l3`).

```
mary_case :: {mary, tom, jim, accident, cold},
 poor/[agent=mary, relevance=l1],
 abandon/[agent=mary,
         coagent=tom/[mother=mary, age=4months],
         relevance=l2],
 find/[agent=jim, object=tom/[state=crying],
         relevance=l1],
 make/[agent=jim, object=accident, relevance=l2],
```

```
injure/[agent=jim, coagent=tom, by=accident,
        relevance=12],
leave/[agent=jim, coagent=tom, relevance=13],
death/[agent=tom, cause=cold, relevance=13]};;
```

The attorneys on both sides interpreted Mary's case according to individual perspectives: one is the responsibility of Mary's actions and the other is that of Jim's. For instance, one attorney reasoned that: "If Mary hadn't abandoned Tom, Tom wouldn't have died. In addition, the cause of Tom's death is not injury but freezing. Therefore there exists a causality between Tom's death and Mary's abandoning."

Another lawyer, however, argued differently: "A crime was committed by Jim, namely, his abandoning Tom. And in addition, Tom's death was indirectly caused by Jim's abandoning Tom. Therefore, there exists a causality between Tom's death and Jim's abandoning."

For a legal precedent, these contradictory claims are documented together with the final verdict from the judge overseeing that precedent. *QUIXOTE* models these arguments with two *case rules* of different interpretations of causality.

```
cr1 :: responsible/[agent=mary,for=death]
        <=
       abandon/[agent=mary,coagent=tom],
       death/[agent=tom,
              cause=abandon/[agent=mary,coagent=tom]];;

cr2 :: responsible/[agent=jim,for=death/[agent=tom]]
        <=
       leave/[agent=jim, coagent=tom],
       death/[agent=tom, cause=leave];;
```

The idea of an *induced rule* is to abstract some of ground terms in local case rules. As an example, when there are several similar accident cases, the attorneys may make the following generalization:

```
ir1 :: responsible/[agent=X, to=Y, for=Inj]
        <=
       Acc/[agent=X],
       Inj/[agent=Y, cause=Acc]
       || {Acc =< accident, Inj=<physical_damage,
       X =< person, Y =< person};;
```

In `ir1`, traffic accident and injury are abstracted to variable `Acc` and `Inj` and subsumed by their super concepts in the legal dictionary.

Legal rules, or statutes, are formal sentences of codes. We provide a penal code in linguistic form (Japanese penal code, article 199): "In case an intentional action of person A causes the death of person B and the action is not presumed to be legal, A is responsible for the crime of homicide."

The *QUIXOTE* representation of this code is:

```
lr1 :: responsible/[agent=A, to=B, for=homicide]
        <=
       Action/[agent=A],
       illegal/[act->Action],
       death/[agent=B, cause->Action],
       || {Action =< intend, A =< person, B =< person};;
```

In the description above, `illegal[agent=A, action = Action]` claims that the action `Action` done by `A`, such as self-defense, is not legal. The statute for the legality of self-defense is described as follows (Japanese penal code, article 38):

```
lr2 :: illegal/[act = Action]
        <=
       Action,
       || {Action =< intend};;
```

The concept of *anchoring* of $\mathcal{SM}$, mentioned in Section 1.3, is realized in *QUIXOTE* by invoking appropriate rules within a case or statute description.

36.2.2 Query Processing

Let us consider Mary's case, where *QUIXOTE* draws several conclusions by making different assumptions. In response to the query:

```
?-responsible/[agent=jim, to=tom, for=homicide].
```

that means "Is Jim responsible to Tom for the crime of homicide?", *QUIXOTE* returns the following:

```
** 2 answers exist **
** Answer 1 **
 IF mary_case:death.cause =< leave THEN
 YES
** Answer 2 **
 IF mary_case:death.cause =< traffic_accident THEN
 YES
```

The first answer is one interpretation of the causality in Mary's case: if the cause of Tom's death is some event under Jim's leaving Tom, then Jim is responsible for the homicide. The latter answer says that Jim is responsible if Tom had been killed by Jim's traffic accident. It happens, however, that the latter does not hold, so that the inquiring agent starts a new query which adds information about the cause of Tom's death.

```
?-mary_case:responsible || {mary_case:death.cause==leave}.
```

In response to this second query, the $\mathcal{QUIXOTE}$ system replies as follows.

```
** 1 answer exists **
** Answer 1 **
 IF mary_case:death.cause == leave THEN YES
```

Thus, we have shown the implementation of our situated inference model in $\mathcal{QUIXOTE}$.

36.3 Conclusion

In this paper, we formalized legal reasoning in terms of $\mathcal{SM}$, where precedent cases and new accidents were regarded as situations, and various kinds of rules as situated inference rules. We also showed that the abstract model was implemented in $\mathcal{QUIXOTE}$ for prototyping. $\mathcal{QUIXOTE}$ could represent context-dependent knowledge and situated inference for knowledge base applications. The ability of $\mathcal{QUIXOTE}$ to model abstract concepts of situation theory in a database environment may pave the way for the knowledge-base (KB) community to tackle concrete, demanding problems, such as building a large scale KB for general linguistic concepts.

References

Ashley, K. 1990. *Modeling legal argument.* MIT Press, Cambridge, MA.

Barwise, J. 1989. *The Situation in Logic.* CSLI Lecture Notes 17.

Barwise, J., and J. Perry. 1983. *Situations and Attitudes.* The MIT Press.

Devlin, K. 1991. *Logic and Information.* Cambridge University Press.

Gardner, A.v.d.L. 1987. *An artificial intelligence approach to legal reasoning.* MIT Press, Cambridge, Massachusetts.

Nitta, K., Y. Ohtake, S. Maeda, M. Ono, H. Ohsaki, and K. Sakane. 1992. HELIC-II: A Legal Reasoning System on the Parallel Inference Machine. In *FGCS '92*, 1115–1124.

Nitta, K., S. Wong, and Y. Ohtake. 1993. A computational model for trial reasoning. In *Proc. 4th Int. Conf. on AI and Law, Amsterdam.*

Rissland, E. L. (ed.). 1991a. Special issue: AI and Legal Reasoning, Part 1. *International Journal of Man-Machine Studies* 34(6).

Rissland, E. L. (ed.). 1991b. Special issue: AI and Legal Reasoning, Part 2. *International Journal of Man-Machine Studies* 35(1).

Sergot, M. 1990. The representation of law in computer programs: a survey and comparison. In *Knowledge Based Systems and Legal Applications*, ed. Bench-Capon. Academic Press.

Wong, S. 1992. A situation-theoretic model for trial reasoning. In *Proc. of the 6th Int. Symp. on legal knowledge and legal reasoning systems, Tokyo*, 32–54. Oct.

Yokota, K., H. Tsuda, Y. Morita, S. Tojo, and H. Yasukawa. 1993. Specific features of a deductive object-oriented database language $\mathcal{QUIXOTE}$. In *Proc. of the workshop on combining declarative and object-oriented databases, ACM SIGMOD, Washington, D.C.* May.

Yokota, K., and H. Yasukawa. 1992. Towards an Integrated Knowledge-Base Management System. In *International Conference on Fifth Generation Computer Systems 1992*, 89–112. Institute for New Generation Computer Technology.

37

An E-type Logic

JAAP VAN DER DOES

Abstract

This short note is on quantification and anaphora. It presents DQL, a dynamic quantifier logic in which some pronouns are treated as E-types. In the literature, the term 'E-type' is used for different proposals. The introduction explains how it is understood here. Next, section 37.1 presents three puzzles which any E-type logic should attempt to solve. Section 37.2 defines the logic DQL that is used in section 37.3 to give solutions to the puzzles posed. The logical development of DQL is left for the full paper.

Introduction

E-types flourish. Roundabout 1980 there were a few publications on this topic starting with the seminal papers of Evans. Cf. Evans 1977, 1980, Cooper 1979, Richards 1984. Ten years later, it surfaced again in Lappin 1989, Heim 1990, Kadmon 1990, Neale 1990, Chierchia 1992, Gawron, Nerbonne, and Peters 1992, Van der Does 1993, 1994, Jackson 1994, and Lappin and Francez 1994. In these papers, the term 'E-type pronoun' is used in different ways. Some authors treat them as terms, others as quantifiers. Some argue for a sloppy interpretation, others for a precise one. This introduction explains how the term is understood here.

Geach claimed that, laziness aside, pronouns come in two sorts: the deictic and the bound ones. Evans (1977) argued against this by hold-

My stay in Moraga was supported by NWO, SIR 11-1211, and by the Research Institute for Language and Speech (OTS, University of Utrecht). This paper is part of the Pionier project 'Reasoning with Uncertainty', NWO 030-22-262. All financial support is kindly acknowledged. I would also like to thank David Beaver, Martin van den Berg, David Israel, Eric Jackson, Michiel van Lambalgen, Kees Vermeulen, Dag Westerståhl and Henk Zeevat for comments and discussions.

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

ing that there is a third variety, characterized as the pronouns which are not c-commanded by their antecedent. He called them E-type pronouns. Sentence (1) is a familiar instance.

(1) Bill owns just two^x sheep, and Harry vaccinates them$_x$
 a. Bill owns just two sheep which Harry vaccinates
 b. Bill owns just two sheep, and Harry vaccinates the sheep owned by Bill

Here and elsewhere, I use the convention that antecedents are indicated by means of a superscript and anaphora by means of a subscript. In (1), the pronoun 'them' has the determiner 'two' as its antecedent. Yet, it is not bound by it. For if it were so bound, the second conjunct of (1) would be part of the determiner's scope. This renders (1) equivalent to (1a), but (1a) is just one of its consequences. Instead, (1) is synonymous with (1b), where the description 'the sheep owned by Bill' comes about by restricting the phrase 'the' to material provided by 'two'. That is, the semantics of (1) is given by (2).

(2) $\textbf{just two}(S, \{d : Ojd\}) \wedge \textbf{pro}(S \cap \{d :$

Here, **pro** is a relation between sets, left unspecified for the moment, whose first argument derives from its antecedent **just two**. In other words, the set $S \cap \{d : Ojd\}$ acts as a *context set* for the quantifier **pro** in the sense of Westerståhl 1984.

Recall that a type $\langle 1, 1\rangle$ determiner $\mathbf{D}$ is usually seen as a functor which assigns to each domain E a relation $\mathbf{D}_E \in \wp(\wp(E) \times \wp(E))$. Natural language determiners put some constraints on their denotations. For instance, they are conservative and extensional:

$$\begin{array}{ll} \text{CONS} & \mathbf{D}_E XY \Leftrightarrow \mathbf{D} X X \cap Y \\ \text{EXT} & \mathbf{D}_E XY \Leftrightarrow \mathbf{D}_{E'} XY, \text{ for } X, Y \subseteq E \subseteq E' \end{array}$$

Westerståhl (1984) made the useful observation that each type $\langle 1, 1\rangle$ determiner $\mathbf{D}_E$ can be contexualized to a set $C \subseteq E$ as follows:

(3) $\mathbf{D}^C_E XY \Leftrightarrow_{\text{df}} \mathbf{D}_E C \cap X, Y$

For conservative and extensional $\mathbf{D}$ this means that one has:

(4) $\mathbf{D}^C_E XY \Leftrightarrow \mathbf{D}_{C \cap X} C \cap X, C \cap X \cap Y$

On this view, different context sets yield different domains of quantification. Accordingly the resolution of E-type anaphora can be seen as a process of domain dynamics. More in particular, the main features of the E-type anaphora formalized in section 37.2 can be summarized by:

- E-type anaphora are context dependent quantifiers;
- The context used in interpreting E-types is a set provided by its antecedent;

- Context change is a process of domain dynamics, which ensures that at each point in a discourse the E-types have resource to the proper context sets.

This picture of anaphor resolution seems attractive for its simplicity. But of course there are several complications which make the general picture more involved, and logically more interesting. They are presented as 'puzzles'.

37.1 Three Puzzles

Each semantics for E-type anaphora should attempt to solve the following three puzzles, ordered by increasing complexity.

37.1.1 First Puzzle: What's in Store?

A pronoun which is anaphorically linked to an antecedent may itself function as antecedent to another anaphor. Formally this means that double indices are not precluded. E.g., in case of 'pro_x^y' the pronoun is linked to an antecedent 'det^x', and will function as antecedent to pronouns 'pro_y'.[1] One may wonder what kind of domain dynamics corresponds to this situation.

In general, 'pro_x^y' uses the context set C_x in its restriction, which is supplied by 'det^x'. So, the context set passed on by 'pro_x^y' can be seen as an update or refinement of C_x. But what about C_x after the update has taken place? Is it still available, or is it used up? The little discourse in (5) shows that it should remain available for further use.

(5) Four^x children play outside. The$_x^y$ girls play marbles.
The$_x^z$ boys play with a hoop. They enjoy themselves.

The domain dynamics of (5) is roughly as follows. The determiner *four* introduces the set of children that play outside. This set is the domain of quantification for the subsequent anaphor sentences: the girls who play marbles are among the children who play outside, and similarly for the boys playing with a hoop. There are several possible antecedents for the pronoun *they*, which differ in availability. Either the pronoun is linked to *four*, or to one of the definite articles. In the first case *they* quantifies over the set of children who play outside. In the other cases there is, say, a restriction to the girls in that set.

It seems natural to model this process by means of files or stores. The first puzzle is therefore: How to formalize the storage mechanism?

37.1.2 Second Puzzle: Dependent Means?

The second puzzle stems from the fact that the domains used in interpreting E-type anaphora may be dependent on each other. Moreover, the dependencies change with the linguistic context in which the pronoun occurs. To

[1] Of course, whenever the antecedent is a non-complex NP it will be a quantifier.

see this, consider a situation in which an antecedent but not its anaphor lies within the scope of a determiner, as in (6).

(6) Every^x farmer who owns just one^y donkey with a^z wooden leg paints it$_z$ grey

a. $[\forall x : Fx \wedge [1y : Dy](Oxy \wedge [\exists z : Lz]Hyz)][\forall z : Lz \wedge H\underline{y}z]Pxz$

b. $[\forall x : Fx \wedge [1y : Dy](Oxy \wedge [\exists z : Lz]Hyz)]([\forall y : Dy \wedge Oxy \wedge \wedge [\exists z : Lz]Hyz][\forall z : Lz \wedge Hyz]Pxz)$

In (6) the quantifier *it* is linked anaphorically to the antecedent *a* in *a wooden leg*. This antecedent introduces the parameterized set *leg of* y. If this set were to restrict the pronoun, the variable y would remain free. Cf. (6a). But (6) means something like (6b), which states that every farmer who owns just one donkey with a wooden leg paints each such leg of every such donkey grey. The same phenomena occurs when anaphoric links cross sentence boundaries, as in (7).

(7) These^x donkeys have just two^y wooden legs. They$_y$ are painted grey.

a. $[\text{these } x : Dx][2y : Ly]Hxy.\ [\forall y : Ly \wedge H\underline{x}y]Gy.$

b. $[\forall x : Dx \wedge [2y : Ly]Hxy][\forall y : Ly \wedge Hxy]Gy$

On the other hand, there are also situations in which the parameters need not be taken care of, since they are already bound by the relevant operators. Cf. the variable x in (6), and in the donkey sentences below.

The second puzzle is therefore: How to account for the dependencies among the domains, and for the ways in which they vary per context?

37.1.3 Third Puzzle: Ambiguities?

The third puzzle concerns the quantifier which a pronoun denotes. Which one could it be? And is there an unambiguous choice over all linguistic contexts? Neale 1990 argues that singular pronouns denote numberless descriptions, given by:

(8) $\mathbf{pro}^C_{\text{sg}}XY \quad \Leftrightarrow \quad C \cap X \subseteq Y \;\&\; |C \cap X| \geq 1$

Cf. also Richards 1984. Neale's admirable defence provides a wide array of examples where numberless descriptions work fine. However, the recent discussion on weak and strong readings of donkey sentences seems to indicate that for singular pronouns no uniform choice of denotation is available. In (9a) *it* should denote the universal quantifier, as in (9b), but in (9c) it should denote the existential one, as in (9d).

(9)

a. Every farmer who owns a donkey beats it

b. $[\forall x : Fx \wedge [\exists y : Dy]Oxy][\forall y : Dy \wedge Oxy]Bxy$

c. Some farmer who owns a donkey beats it

d. $[\exists x : Fx \wedge [\exists y : Dy]Oxy][\exists y : Dy \wedge Oxy]Bxy$

Similar phenomena occur within the neighbourhood of connectives. E.g., the pronouns in (10a) have universal force, and those in (10b) existential force.

(10)
a. If a farmer owns a donkey he beats it
b. A farmer owns a donkey and he beats it

There also appear to be mixed cases such as in (11).

(11)
a. Every farmer who owns a^y whip and a^x stallion uses it_y to lash him_x
b. $[\forall x : Fx \wedge [\exists y : Wy](Oxy) \wedge [\exists z : Sz](Oxz)]([\exists y : Wy \wedge Oxy] [\forall z : Sz \wedge Oxz]Lxzy)$

Clear intuitions are lacking, but (11a) seems to be true even if the farmers use but one of whip to lash all of their stallions (cf. 11b).

The third puzzle is therefore: How to account for these differences in quantificational force in a principled way?

This ends the list of puzzles. I now define an E-type logic which aims to solve all of them.

37.2 Dynamic Quantifier Logic

This section defines DQL, which is short for: dynamic quantifier logic. DQL is best thought of as a standard logic, where all NPs (definite, indefinite or otherwise) are interpreted as quantifiers. Some of these quantifiers depend on the context generated by a text (pronouns are a prime example). To model this, DQL comes with a separate storage mechanism.

37.2.1 The Language

We first introduce a 'core' language which focuses on the essential property of the system; namely, that determiners may either introduce new information or update already available information. We work under the assumption that all determiners are used in both ways, indicated by means of an indexing system. This is too liberal. But I think of the system as providing a logical space of which the 'real' part may by singled out by means of constraints (cf. Van Deemter 1991, among others).

Let $\mathcal{L}$ be a language which has individual constants c, d, ...; identity '=' besides relation signs R^n, S^m, ... of the indicated arity, and two place determiner signs 'all,' 'some,' D, D′, ... Besides, there is an infinite supply of variables. The letters φ, ψ, ... denote formulas.

Definition 1 *The set of* formulas *is the smallest set satisfying:*

i) If $x_1 \ldots x_n$ are variables and R is an n-place relation sign, then $Rx_1 \ldots x_n$ is an atomic formula.

ii) If x and y are variables, then $x = y$ is an atomic formula.

iii) *If φ and ψ are formulas, then so are $\neg\varphi$, and $\varphi \wedge \psi$.*

iv) *If x, y are variables,* D *a determiner sign, and φ and ψ are formulas, then $[\mathrm{D}_y x : \varphi]\psi$ is a formula.*

A main feature of the language is its treatment of quantificational expressions. In DQL, as in DRT and DPL, variables play a double role: they are used in the binding mechanism, and they indicate anaphor-antecedent dependencies. In the formal language this shows up in the fact that determiners come with two variables: a subscripted and a 'normal' one. The determiner binds the normal variable. The subscripted variable indicates the context set of the determiner, i.e., the set introduced by the antecedent which binds that variable (if any). All definitions below assume that each occurrence of a determiner comes with a unique normal variable. Other determiners use this variable only as a subscript. Now we formalize the storage mechanism used to provide context sets.

37.2.2 Storage

In interpreting E-type anaphora it is crucial that a text generates context sets to supply their restriction. Relative to a model, the context sets used in DQL are all definable. Roughly speaking, they consist of the objects which satisfy the formula introduced by the antecedent. Note, however, that the puzzle of dependent means indicates that we cannot specify the defining formula once and for all. It varies with the dependencies among the context sets, and these, in turn, vary with the structural position of the anaphoric element. More in particular, which dependencies are in force depends on the variable-binding operators in whose scope the anaphor occurs. This means that the store should provide two things: (i) the building blocks of the defining formulas, and (ii) a dependency relation which is used in their construction. The formal objects which we employ for this purpose are partial functions from variables to formulas; objects which are much like the files in Heim (1982,1983).[2] In terms of such files, the *storage effect* of a formula can be specified inductively.

Definition 2 *With each formula φ we assign a function $(\!|\varphi|\!)$ from files to files. Using postfix notation, we have:*

i) $\mathbf{f}(\!|Rx_1 \ldots x_n|\!) = \mathbf{f}$

ii) $\mathbf{f}(\!|x = y|\!) = \mathbf{f}$

iii) $\mathbf{f}(\!|\neg\varphi|\!) = \mathbf{f}$

iv) $\mathbf{f}(\!|\varphi \wedge \psi|\!) = \mathbf{f}(\!|\varphi|\!) \cup \mathbf{f}(\!|\psi|\!)$.

v) $\mathbf{f}(\!|[\mathrm{D}_y x : \varphi]\psi|\!) = \{\langle x, [x/y]\mathbf{f}(y) \wedge \varphi \wedge \psi\rangle\} \cup \mathbf{f}(\!|\varphi|\!) \cup \mathbf{f}(\!|\psi|\!)$

By definition '$[x/y](\mathbf{f}(y))\wedge$' is the empty string if $y \notin \mathrm{DOM}(\mathbf{f})$.

[2]Below I use a slightly extended notion of file.

Definition 2 uses a constructive view on the increase of information. Rather than eliminating possibilities, as is done in DPL and other systems, stores are 'structured objects' which are created as we go along. At this point we only distinguish between operators which either do or do not contribute to the content of the store (determiners vs. the rest). In other respects, the store simply collects all relevant information given by the preceding discourse. This information will be used in different ways in different linguistic contexts. Before turning to examples we give some formal comments.

37.2.2.1 Non-quantificational Sentences

Conjunctions just add the information provided by their subformulas. The order in which the formulas store the information is immaterial. Order only plays a role in the interpretation of formulas (cf. section 37.2.3). Negation is treated as creating an anaphoric island, which disallows links with anaphora to the right of its scope. An alternative would be to define:

$$\mathbf{f}(\llbracket\neg\varphi\rrbracket) = \mathbf{f}(\llbracket\varphi\rrbracket)$$

This form of negation does not block any anaphoric link. DPL and DMG allow corresponding options. Dekker (1993, ch. 2) shows that these have to be refined. It is not clear whether his observations apply to DQL, but for present purposes the above definition will do.

37.2.2.2 Quantificational Sentences

Quantificational sentences store a formula which defines the context set used to interpret E-type pronouns.[3] In doing so, we have to ensure that the functionality of files is preserved. But this is taken care of by the fact that each determiner occurrence comes with a unique normal variable.

The treatment of quantifiers in DQL differs from that of DRT and DPL, which hold that quantifiers other than indefinites induce anaphoric islands (they are 'tests'). For plural anaphora, not handled by the initial versions of these systems, we have seen several examples showing that this is too strict. But singular pronouns also require more freedom.

(12)
a. Every farmer owns a donkey. *He beats it.
b. *If a farmer owns every donkey he beats it.
c. If a farmer gives every donkey a stroke it isn't painful.
d. Just one farmer owns a donkey. It is mouse-grey.

In (12a,b) the singular pronouns cannot be anaphorically linked to the quantifier 'every'. Still, the quantifier should not preclude all anaphoric links. For instance, 'every' should have wide scope over 'a' in (12c). But

[3] With a view to noun anaphora, one may consider to make a file a function from variables to pairs of formulas. This would enable to choose between a noun anaphor (projection on the first argument) or an E-type anaphor (conjunction of the elements of the pair). I use the simpler format, since I do not consider noun anaphora.

the singular pronoun 'it' may depend on the indefinite. The same holds for (12d) even if the quantifier 'just one' has wide scope. Whether or not quantifiers create anaphoric islands, the distribution remains unexplained. Here I presume that anaphoric islands are the exception rather than the rule.

37.2.2.3 Dependency

We required that the storage mechanism should come with a dependency relation to obtain the appropriate context sets. For DQL this is defined as follows.

Definition 3 *Let* $\mathbf{f}$ *be a partial function from variables to formulas. For* $x, y \in \text{DOM}(\mathbf{f})$ *we say that* y depends on x *(in* $\mathbf{f}$*)—notation:* $x <_{\mathbf{f}} y$*—iff: there are* $z_1 \ldots z_n \in \text{DOM}(\mathbf{f})$ *with* $z_n = x$, $y = z_1$ *and* $z_{i+1} \in \text{FV}(\mathbf{f}(z_i))$.

By definition, the relation $<_{\mathbf{f}}$ is transitive. In general it need not be irreflexive. However, we restrict attention to formulas in which each determiner occurrence has a unique normal variable which does not bind vacuously. Then, the files considered have the property: If $x \in \text{FV}(\mathbf{f}(y))$, then $\text{FV}(\mathbf{f}(x)) \subset \text{FV}(\mathbf{f}(y))$ and $y \notin \text{FV}(\mathbf{f}(x))$. This property ensures irreflexivity after all. The relations $<_{\mathbf{f}}$ formalize the dependency among the domains defined by the stored formulas. As we shall see, they are crucial for a solution to the puzzle of dependent means.

The next section defines the interpretation of formulas within context. First, it introduces a subpart of DQL without singular pronouns. Singular pronouns receive special attention in section 37.3.3.

37.2.3 Interpretation in Context

The semantics of DQL uses standard models. A model $\mathcal{M}$ is a pair $\langle D, ^* \rangle$ consisting of a non-empty domain D and an interpretation function *. The interpretation function * assigns elements of D to individual constants, n-place relations over D to n-place relation signs, and two-place relations between sets of D to determiner signs. E.g., it assigns the inclusion relation to 'all', the relation of non-empty overlap to 'some', and so on. Often I write, say, **all** instead of (all)*. The standard semantics is supplied with a storage mechanism. Definition 4 uses several operations on files, which are explained below.

Definition 4 *Let* $\mathcal{M} = \langle D, ^* \rangle$ *be a model,* $\mathbf{f}$ *a file, and* $\mathbf{a}$ *an assignment for* $\mathcal{M}$*. The* truth of χ in $\mathcal{M}$ with respect to [**a**;**f**] *– notation:* $\mathcal{M} \models \chi$ [**a**;**f**], *– is defined by:*

i) $\mathcal{M} \models Rx_1 \ldots x_n$ $[\mathbf{a}; \mathbf{f}]$ *iff:* $\langle \mathbf{a}(x_1), \ldots, \mathbf{a}(x_n) \rangle \in R^*$
ii) $\mathcal{M} \models x = y$ $[\mathbf{a}; \mathbf{f}]$ *iff:* $\mathbf{a}(x) = \mathbf{a}(y)$
iii) $\mathcal{M} \models \neg\varphi$ $[\mathbf{a}; \mathbf{f}]$ *iff:* $\mathcal{M} \not\models \varphi$ $[\mathbf{a}; \mathbf{f}]$

iv) $\mathcal{M} \models \varphi \wedge \psi$ $[\mathbf{a};\mathbf{f}]$ *iff:* $\mathcal{M} \models \varphi$ $[\mathbf{a};\mathbf{f}]$ *and* $\mathcal{M} \models \psi$ $[\mathbf{a};\mathbf{f}(\!|\varphi|\!)]$

v) $\mathcal{M} \models [\mathrm{D}_y x : \varphi]\psi$ $[\mathbf{a};\mathbf{f}]$ *iff* $\mathbf{D}(C \cap X, Y)$ *with:*

- $C = \widehat{y}.[\![\mathbf{f}(y)]\!]_{\mathbf{a}\,\mathbf{f}}$
- $X = \widehat{x}.[\![\varphi]\!]_{\mathbf{a},([x/y]\mathbf{f})^x}$
- $Y = \widehat{x}.[\![\psi]\!]_{\mathbf{a},([x/y](\mathbf{f}(\!|\varphi|\!)))^x}$

By definition, $\widehat{x}.[\![\varphi]\!]_{\mathbf{a},\mathbf{f}}$ *is the set* $\{d \in D \mid \mathcal{M} \models \varphi\ [\mathbf{a}[d/x];\mathbf{f}]\}$.

In DQL, plural definites are just a particular kind of determiner. Formally, a plural pronoun such as *they* is treated as a definite with empty descriptive content:

(13) $[\mathrm{pro}_y\, x]\varphi \equiv [\mathrm{all}_y\, x : \top]\varphi$

Cardinality restrictions on the first argument are ignored.

Despite the partiality of the contexts the semantics is total. Indeed, the semantics is standard except for the domain dynamics reflected in the operations on files.

(14) a. $\mathbf{f}^x = \{\langle x, \surd\rangle\}$

b. $[x/y]\mathbf{f} = \{[x/y]\langle u, \mathbf{f}(u)\rangle \mid u \in \mathrm{DOM}(\mathbf{f})\}$

$$[x/y]\langle u, \varphi\rangle = \begin{cases} \langle x, [x/y]\varphi\rangle & \textit{if } y \equiv u \\ \langle u, [y/x]\varphi\rangle & \textit{else} \end{cases}$$

The operation $\mathbf{f}^x$ marks a variable as bound (much as in Pagin and Westerståhl 1993), and $[x/y]\mathbf{f}$ renames a variable in a file. These operations are used to formalize the claim that the scope relations between anaphor and antecedent are crucial in determining the anaphor's context set. For example, pronouns with a bound context variable should function as bound variables. In order to show how the operations achieve this, I first discuss the three kinds of anaphora discerned in DQL.

37.2.3.1 Three Kinds of Anaphora

In terms of the features 'to be bound' (BND) and 'to have an antecedent' (ANT), DQL discerns deictic, bound, and E-type anaphora. Since in the present set up BND and ANT are incompatible, this means that all logical possibilities along the two dimensions are realized.

	BND	ANT
*	+	+
bound	+	–
E-type	–	+
deictic	–	–

The formal content of the features is as follows.

Definition 5 *A determiner* $D_y x$ *occurs* bound *iff it occurs within the re-*

striction or scope of a determiner $D'y$. It has an antecedent *if* $\mathbf{f}(y)$ *is a formula as soon as* $D_y x$ *is interpreted.*

It is logically impossible for a determiner to be bound and have an antecedent at the same time. An antecedent provides an E-type with a context set that is specified in terms of the material which occurs within its scope. This material is stored under its normal variable. On pain of circularity an antecedent of an E-type anaphor cannot at the same time bind it. Otherwise the context set introduced by the antecedent would have to be specified in terms of material which presumes it.

Some examples are in (15a–c), where the character of the pronoun is stated in the rightmost column.

(15) A man walks and he talks.

a.	$[\exists x : \mathrm{M}x](\mathrm{W}x) \wedge [\mathrm{pro}_z y](\mathrm{T}y)$	*deictic*
b.	$[\exists x : \mathrm{M}x](\mathrm{W}x \wedge [\mathrm{pro}_x y]\mathrm{T}y)$	*bound*
c.	$[\exists x : \mathrm{M}x](\mathrm{W}x) \wedge [\mathrm{pro}_x y](\mathrm{T}y)$	*E-type*

Here and in what follows it is understood that determiners which lack a context variable have a covert one without an antecedent. The semantics ensures that such determiners are contextually unrestricted. In (15) the normal variables are unique per occurrence of a quantifier, therefore (15b) uses x for *an*, and y for *pro*. Nevertheless, due to the fact that the semantics keeps track of the status of variables, the semantics can turn it into a bound variable. To this end we should classify determiners as those which can and those which cannot be bound (cf. DRT, which does so by means of open and closed expressions). Here we could use two sorts of context variable: x and $\overline{x}$. The context variables x can have the determiner $D_y x$ as their antecedent or can be bound by it, while context variables $\overline{x}$ can only have $D_y x$ as an antecedent. When bound, the bindable determiners should function as bound pronouns. But the determiners which cannot be bound should then have their standard, contextually unrestricted meaning. Since files record whether or not a variable is bound, we can discern these cases formally. E.g., for the case that $\mathbf{f}(y) = \surd$ we define:

a. $\mathcal{M} \models [Pro_y x]\varphi\ [\mathbf{a};\mathbf{f}]$ iff $\mathcal{M} \models [y/x]\varphi\ [\mathbf{a};\mathbf{f}]$

b. $\mathcal{M} \models [D_{\overline{y}} x : \varphi]\psi\ [\mathbf{a};\mathbf{f}]$ iff $\mathbf{D}(X, Y)$ with:

- $X = \widehat{x}.[\![\varphi]\!]_{\mathbf{a};([x/y]\mathbf{f})^x}$
- $Y = \widehat{x}.[\![\psi]\!]_{\mathbf{a};([x/y](\mathbf{f}(\![\varphi]\!)))^x}$

On the other hand, in case $D_y\, x$ is an E-type anaphor the determiner should be restricted to the set $\lambda x.[x/y]\mathbf{f}(y)$. But what is more, the substitution $[x/y]$ should be used for the entire file. For whenever $D_y\, x$ has a determiner within its scope whose context variable z depends on y, then x should vary over $\mathbf{f}(z)$ as well. Cf. (21) below. All in all this means that we should use

$[x/y]\mathbf{f}$. The next section, which has solutions to the three puzzles, has some examples which should clarify these remarks.

37.3 Solutions

37.3.1 First Solution: Files as Stores

This section continues the discussion of (5). Let us first compute the storage effects of its formalization. We start ignorant with $\emptyset$, the empty file which is undefined for all variables. This file updated with the sentence *Four*^{x} *children play outside* yields the file $\mathbf{f}_1$ which is just defined for x and has $\mathbf{f}_1(x) \equiv Cx \wedge Px$. The file $\mathbf{f}_1$ updated with the sentence *The*^y_x *girls play marbles* produces $\mathbf{f}_2$ which is defined for x and y, and:

$$\begin{array}{llcl} (16) & \mathbf{f}_2(x) & \equiv & \mathbf{f}_1(x) \\ & \mathbf{f}_2(y) & \equiv & [y/x](\mathbf{f}(x)) \wedge Gy \wedge My \\ & & \equiv & Cy \wedge Py \wedge Gy \wedge My \end{array}$$

The update of $\mathbf{f}_2$ with the sentence *The*^z_x *boys play with a hoop* yields $\mathbf{f}_3$ which is defined for x, y, and z, and:

$$\begin{array}{llcl} (17) & \mathbf{f}_3(x) & \equiv & \mathbf{f}_2(x) \\ & \mathbf{f}_3(y) & \equiv & \mathbf{f}_2(y) \\ & \mathbf{f}_3(z) & \equiv & [z/x](\mathbf{f}(x)) \wedge Bz \wedge Hz \\ & & \equiv & Cz \wedge Pz \wedge Bz \wedge Hz \end{array}$$

Notice that the use of subscripted variables enables us to update information without loosing it. We just update a substitution instance of the old information. Also, in interpreting the anaphoric elements these variables are bound by means of abstraction. Since the identity of bound variables does not matter, we get the desired set in each case.

It remains to show that the truth conditions of (5) are as required. To do so, we consider texts to be conjunctive in nature: the semantics of '.' is that of '$\wedge$'. The first sentence of (5), *Four*^{x} *children play outside*, is non-anaphoric and retains its standard meaning. It is true iff:

$$|[\![C]\!] \cap [\![P]\!]| = 4$$

The sentence *The*^y_x *girls play marbles* depends on the context set introduced by the first. It is true iff [the $y : Cy \wedge Py \wedge Gy]My$ iff:

$$[\![C]\!] \cap [\![P]\!] \cap [\![G]\!] \subseteq [\![M]\!]$$

Similarly, *The*^z_x *boys play with a hoop* is true iff [the $z : Cz \wedge Pz \wedge Bz]Hz$ iff:

$$[\![C]\!] \cap [\![P]\!] \cap [\![B]\!] \subseteq [\![H]\!]$$

Finally, relative to $\mathbf{f}_3$ the sentence *They enjoy themselves* may choose from three context sets. Intuitively, (x) the set of children that play outside, (y)

the set of girls that play marbles outside, and (z) the set of boys that play with a hoop outside. So it will be true iff one of (18x–z) is true.

(18) (x) $[\![C]\!] \cap [\![P]\!] \subseteq [\![E]\!]$
(y) $[\![C]\!] \cap [\![P]\!] \cap [\![G]\!] \cap [\![M]\!] \subseteq [\![E]\!]$
(z) $[\![C]\!] \cap [\![P]\!] \cap [\![B]\!] \cap [\![H]\!] \subseteq [\![E]\!]$

This is just as we want it to be.

37.3.2 Second Solution: Dependency Relation

The problem with dependent means is that they vary with the linguistic context in which the pronoun occurs. But in each instance the logic should take care of itself. To see how DQL fares in this respect, we first consider (19) once more.

(19)
a. These^x donkeys have just two^y wooden legs. They$_y$ are painted grey.
b. [these $x : Dx][2y : Ly]Hxy$. $[\forall y : Ly \wedge H\underline{xy}]Gy$.

Note that the antecedent of 'they' introduces the parameterized set *wooden leg of* x. But if this set were to restrict the pronoun, the parameter x would remain free.

There are two ways in which this problem can be solved. They correspond to the logically equivalent formalizations (20a,b).

(20) a. $[\forall x : Dx \wedge [2yLy]Hxy][\forall y : Ly \wedge Hxy]Gy$
b. $[\forall y : [\exists x : Dx \wedge [2y : Ly]Hxy](Ly \wedge Hxy)]Gy$

The first solution would be to say that the logical form of the anaphor sentence has a covert pronominal element *pro* x, which has scope over the entire sentence. This makes the anaphor sentence mean (20a). The idea would be that the quantificational structure of the antecedent and the anaphor sentence must be strictly parallel to each other. Both should contain two quantifiers, and the scope relation between them should be the same as that between their anaphora. This would accord nicely with similar parallelism phenomena in case of VP anaphora. Cf. Van den Berg, Prüst, and Scha 1994, among many other papers. The second solution holds that the antecedent 'two wooden legs' does not introduce a set but a set of sets; namely the set *legs of* x for each donkey x which has two wooden legs. Then, there is no covert pronoun. Instead, the pronoun is restricted by the union of this family of sets. Formally, the meaning of the anaphor sentence would be given by (20b). Notice that since x does not occur free in Gy, (20a,b) are indeed equivalent.[4]

[4]In the final version of this paper I show how the dependency relation is used to give the general form of the second strategy. Cf. also Van der Does 1994.

Now we consider a second example, which is slightly more involved. It shows why definition 4 requires the substitution on files within the scope of an antecedent.

(21) a. These^x donkeys have two^y wooden legs. They^u_x ignore them^v_y.
b. [these $x : Dx][2y : Ly]Hxy$. [they$_x u$][them$_y v]Iuv$.

The anaphor sentence means: each donkey who has two wooden legs ignores the wooden legs *it* has. We therefore have to make sure that *it* is interpreted as a bound variable which runs over the appropriate wooden legs. Since a variable occurs only once as the normal variable of a determiner this does not happen automatically. It requires its context variable to be renamed by its normal variable.

Observe that the complication would be absent if the normal variable of a determiner were used both for binding and to indicate its context set. But then the normal variable of a determiner could not be unique per occurrence. Recall that I opted for this uniqueness to solve the first puzzle. The storage effect of a determiner is named by its normal variable, and we do not want two entries to have the same name. Although the information associated with a variable may be an update of older information, the old information should remain available. Hence, the update is stored under a new name. As a consequence, we cannot use the material introduced by an antecedent to restrict its anaphora. It can only be so used after renaming the context variables in store to the variable quantified over by the determiner. That is why (i) substitution takes place within the restriction of a determiner, and (ii) this substitution is passed on to its scope.

A worked example should make these remarks clear. Starting from $\emptyset$, the antecedent sentence induces the context $\mathbf{f}$ with:

(22)
$$\begin{aligned} \mathbf{f}(x) &\equiv Dx \wedge [2y : Ly]Hxy \\ \mathbf{f}(y) &\equiv Ly \wedge Hxy \end{aligned}$$

This means that 'they' in the anaphor sentence is correctly restricted to the set (23a) below. But the renamed variables should also be passed to the elements within the scope of 'they'. Otherwise 'them' would be interpreted relative to $\mathbf{f}$ as well, and its restriction would be the set (23b).

(23)
a. $\widehat{u}.[\![[u/x]\mathbf{f}(x)]\!]_{\mathbf{a},\mathbf{f}} \quad (= \quad \widehat{u}.[\![Du \wedge [2y : Ly]Huy]\!]_{\mathbf{a},\mathbf{f}})$
b. $\widehat{v}.[\![[v/y]\mathbf{f}(y)]\!]_{\mathbf{a},\mathbf{f}} \quad (= \quad \widehat{v}.[\![Lv \wedge Hxv]\!]_{\mathbf{a},\mathbf{f}})$

Now, x remains free since 'them' binds v. In case $[u/x]\mathbf{f}$ is used, this does not happen: $[u/x]\mathbf{f}(y) \equiv Ly \wedge Huy$. In sum, we see that the anaphor sentence in (21) gets the meaning:

$$[\forall u : Du \wedge [2y : Ly]Huy][\forall v : Lv \wedge Huv]Iuv$$

as required. This leaves the puzzle of the weak and strong donkeys, which is our final topic.

37.3.3 Third Solution: Quantificational Contexts

The ambiguity puzzle concerned the semantics of singular pronouns. If some singular pronouns are contextually restricted quantifiers, as an E-type analysis would have it, it appears as if their denotation cannot be chosen uniformly. In some contexts they should denote existential quantifiers, in other contexts universal ones. The weak and strong readings of donkey sentences are a case in point. In this section I sketch a solution to this puzzle, which uses a variant of choice functions. Cf. Chierchia 1992, Gawron, Nerbonne, and Peters 1992, Van der Does 1993, 1994, Lappin and Francez 1994. Let a model $\mathcal{M}$ be of the form $\langle D, E, ^* \rangle$. Here $D = E \cup \{\bullet\}$, with $\bullet \notin E$ the null object which is the value of a choice from the empty set. The null object is disallowed to occur in the extension of relations (an alternative would be to use a partial semantics).

Definition 6 *A* choice function $\mathbf{h}$ *for a set* X *is a function* $\mathbf{h} : \wp(X) \longrightarrow X \cup \{\bullet\}$ *with:*

$$\begin{cases} \mathbf{h}(Y) \in Y & \text{if } Y \neq \emptyset \\ \mathbf{h}(Y) = \bullet & \text{otherwise} \end{cases}$$

A generalized assignment *for a model* $\mathcal{M} = \langle D, E, ^* \rangle$ *is a function* $\mathbf{g}$ *which assigns to each variable a choice function for* E.

A generalized assignment essentially encodes a set with an assignment for each non-empty subset of E. In particular, $\lambda x \in \text{VAR}.\mathbf{g}(x)(E)$ is an assignment for E, and we write $\mathbf{g}(x)$ for $\mathbf{g}(x)(E)$. Moreover, $\mathbf{h}[x]\mathbf{g}$ iff $\mathbf{h}$ differs at most from $\mathbf{g}$ in its value $\mathbf{h}(x)(Y)$, for all non-empty $Y \subseteq E$. Except for these conventions, the definition of truth in context remains unaltered. A singular pronoun is now interpreted relative to a generalized assignment and a file, as follows:

(24) $\mathcal{M} \models [\text{pro}_y^{\text{sg}}\, x]\varphi\ [\mathbf{g}; \mathbf{f}] \Leftrightarrow \mathbf{g}(y)(\hat{y}.[\![\mathbf{f}(y)]\!]_{\mathbf{g};\mathbf{f}}) \in \hat{x}.[\![\varphi]\!]_{\mathbf{g};([x/y]\mathbf{f})^x}$

For example, '*A man walks in the park. He whistles.*' is true iff a choice from the set of men who walk in the park whistles. This semantics of singular pronouns allows DQL to attribute apparent differences in quantificational force not to the pronouns but rather to the quantificational elements in the context of which they occur. To make this precise, let $\mathbf{h}[X]\mathbf{g}$ mean that $\mathbf{h}[x]\mathbf{g}$ for all $x \in X$. Now define:

$$\mathcal{M} \models [D_y^{\text{wr}}\, x : \varphi]\psi\ [\mathbf{g}; \mathbf{f}] \text{ iff } \mathbf{D}(C \cap X, Y)$$

with C and X as before but with Y changed to:

$$Y = \{d \in D \mid \exists \mathbf{h}[\text{DOM}(\emptyset(\![\varphi]\!)))]\mathbf{g} : \mathcal{M} \models \varphi\ [\mathbf{h}[d/x]; ([x/y](\mathbf{f}(\![\varphi]\!)))^x]\}$$

Similarly, the strong reading of a determiner, $[D_y^{\mathrm{sr}}\, x : \varphi]\psi$ is obtained by means of a universal quantification over generalized assignments in Y.

The most prominent characteristics of the weak and strong readings are (i) that the quantification in the second argument is restricted to the new antecedents introduced by its first (cf., DRT and DPL), and (ii) that they are no anaphoric islands. For reasons of time and space, I have to leave it to the reader to verify that for suitable choices of φ and ψ these readings give exactly the same truth conditions to donkey sentences as DRT and DPL do. They also enjoy analogues of Kanazawa's monotonicity results, which allows a suitable choice in case of doubly monotone determiners. Cf. Van der Does 1994 for more details.

References

Chierchia, Gennaro. 1992. Anaphora and dynamic binding. *Linguistics and Philosophy* 15:111–183.

Cooper, Robin. 1979. The Interpretation of Pronouns. In *Selections from the Third Groningen Round Table*, ed. F. Heny and H. Schnelle. 61–92. Syntax and Semantics (vol. 10). New York: Academic Press.

Dekker, Paul. 1993. *Transsentential Meditations. Ups and Downs in Dynamic Semantics*. Ph.D. Thesis. Amsterdam: University of Amsterdam.

Evans, Gareth. 1977. Pronouns, Quantifiers and Relative Clauses I & II. *Canadian Journal of Philosophy* 7:467–536, 777–797.

Groenendijk, Jeroen, and Martin Stokhof. 1991. Dynamic Predicate Logic. *Linguistics and Philosophy* 14:39–100.

Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases.* Ph.D. Thesis. Amherst: University of Massachusetts.

Heim, Irene. 1990. E-type Pronouns and Donkey Anaphora. *Linguistics and Philosophy* 13:137–177.

Jackson, Eric. 1994. *Donkey Sentences, Description and Quantification.* Unpublished manuscript. Stanford: Stanford University.

Kamp, Hans, and Uwe Reyle. 1993. *From Discourse to Logic.* Dordrecht: Reidel.

Kanazawa, Makoto. 1993. Weak vs. Strong Readings of Donkey Sentences and Monotonicity Inference in a Dynamic Setting. *Linguistics and Philosophy* 17:109–158.

Lappin, Shalom, and Nissim Francez. To appear. E-type Pronouns, I-Sums, and Donkey Anaphora. *Linguistics and Philosophy.* Also Report LCL 94–4, Technion - Israel Institute of Technology, Isreal.

Neale, Stephen. 1990. *Descriptions.* Cambridge (Mass.): The MIT Press.

Pagin, Peter, and Dag Westerståhl. 1993. Predicate Logic with Flexibly Binding Operators and Natural Language Semantics. *Journal of Logic, Language, and Information* 2:89–128.

Richards, Barry. 1984. On Interpreting Pronouns. *Linguistics and Philosophy* 7:287–324.

van Deemter, Kees. 1991. *On the Composition of Meaning*. Ph.D. Thesis. Amsterdam: University of Amsterdam.

van der Does, Jaap. 1994. Formalizing E-type anaphora. In *Proceedings 9th Amsterdam Colloquium*, ed. Paul Dekker and Martin Stokhof. Amsterdam: University of Amsterdam.

Westerståhl, Dag. 1984. Determiners and Context Sets. In *Generalized Quantifiers in Natural Language*, ed. Johan van Benthem and Alice ter Meulen. 45–71. GRASS series. Dordrecht: Foris.

38

From Utterances to Situations: Parsing with Constructions in Small Domains

WLODEK ZADROZNY

Introduction

We describe data structures and algorithms for computing parameters of situations from NL utterances in conversation-for-action dialogs. We observe that any type, i.e. any member of the domain ontology, can be a parameter of a situation, and that the goal of a conversation-for-action dialog is to establish values of the relevant parameters. Regarding the data structures, we encode linguistic information in *constructions*, which simultaneously encode syntactic, semantic and pragmatic information about how the language is used. With respect to the algorithms, we use in the process of parsing both linguistic and non-linguistic knowledge, the latter comprising of information about the domain of discourse and about the particular application. We argue that this approach is very suitable for NL understanding in small domains, and illustrate it with examples of understanding of sentences with prepositional phrases.

The Problem

We are working on building systems capable of conversation about actions in simple domains. For example, we have built a natural language interface, MINCAL, to an on-line calendar (Zadrozny et al. 1994). In this system the user can schedule and cancel appointments by talking to the computer or typing phrases. Currently, we are extending the natural language understanding capabilities of the system to other tasks such as room scheduling

Logic, Language and Computation.
Jerry Seligmann and Dag Westerståhl, eds.

and email, and investigating the applicability of our techniques to other domains.

The user and computer converse by specifying parameters of their current situation, typically the type of the action and its parameters. Very often these parameters are given in sentences (or fragments) that could be structurally and/or semantically ambiguous. Removing the ambiguity is thus a necessary condition for capturing those parameters.

Example. Consider the phrase *from 5 to 7*. Even in the restricted domain of calendar scheduling it can mean either of the following: *range*([5, 7]), *beginning*(6: 55), *beginning*(5) & *end*(7), or *old_time*(5) & *new_time*(7). Furthermore, *5* and *7* may also refer to room numbers, or the number of people expected in a room, adding yet another set of meanings. Also notice that this phrase can appear either separately or as a part of a longer utterance, for instance

Schedule a meeting on parsing with Bob in the cafeteria from 5 to 7 and that getting this phrase in the longer sentence is a mixed blessing: additionally, we seem to face the problem of "PP attachment"; but, intuitively, we also can eliminate most of the possible meanings of the phrase *from 5 to 7*

Thus, the problem of computing parameters of situations is a part of the problem of computing single meanings of utterances In particular, of computing single meanings of utterances containing prepositional phrases. Notice that although natural languages in general are ambiguous, it does not necessarily follows that this problem is difficult in small domains we are interested it; for instance, we could simply avoid ambiguities by building a semantic grammar (cf. Section 2) But we want to have a grammar that can also be used for other tasks, so the problem cannot be solved that way. Also, asking a user "did you mean x?" each time an ambiguity is encountered does not make sense. Thus another method is needed.

Our solution is a combination of three elements: (1) limiting structural ambiguities by using a grammar of constructions, where forms, meanings and contexts are integrated in one data structure; (2) using background knowledge during parsing; (3) using discourse context during parsing and semantic interpretation.

Sketch of the Disambiguation Method

Let us begin by explaining what we are *not* doing. (1) We avoid the PP-attachment problem by directly computing the meanings of adjuncts (including PPs), without first trying to establish what is attached to what. To be sure, sometimes we are left with multiple semantic analyses, but without the problem of attaching the PPs. (2) We are not using a semantic grammar; that is, a grammar which has built in semantic constraints for a particular domain, such as calendars (with rules like `VP` $\rightarrow$ `schedule_Verb meeting_NP`) or carpentry (cf. Gazdar and Mellish 1989, pp.306ff).

These two points are easy to justify. Point (1) is obvious — it is better to avoid a problem instead of trying to solve it. Secondly, using a semantic grammar would contradict our aim of building a grammar that can be used for other tasks.

We now can briefly explain the method; these explanations will then be elaborated in the subsequent sections. We have to describe (A) the data structures and (B) the algorithm.

The Data Structures

Since language understanding is impossible without background knowledge, we have to describe not only the way we represent general linguistic information, but also how background knowledge is represented. Let us start with the latter. At this point it is enough to know that (i) we represent background knowledge as sets of theories, i.e. not one flat knowledge base (cf. Zadrozny 1994b), and that we divide it into the knowledge of domain (e.g. facts about months, dates, etc.), the knowledge of the application (e.g. about obligatory and optional parameters of meetings in *xdiary*), linguistic facts about the domain (e.g. that dates do not modify *np(person)*), and linguistic facts about the application (e.g. that "places" (*np(place)* or *pp(in, place)*) do not modify *np(person)*, which is not true in general).

We use *constructions* to represent general linguistic knowledge. A *construction* is a data structure consisting of a description of context in which it applies, its form, and its message (meaning). It combines in one description syntactic, semantic and pragmatic features. Constructions range from words to discourse; they include both the standard constructions of type *S* $\rightarrow$ *NP VP* discussed by generative grammarians, and the not so standard constructions like *The X-er, the Y-er* or reduplication analyzed by Fillmore, Kay, Manaster Ramer and others.

A construction is given by the matrix:

$$\begin{bmatrix} \mathbf{N} : name\_of\_construction \\ \begin{bmatrix} \mathbf{C} & : & context \\ \mathbf{V} & : & structure \\ \mathbf{M} & : & message \end{bmatrix} \end{bmatrix}$$

The *structure* (or *vehicle*) **V** consists of formulas describing presence (or perhaps absence) of certain *taxemes*, or features of form, within the structure of the construction. Such a structure is given by a list of subconstructions and the way they have been put together (in all our examples this is concatenation, but there are other possibilities, e.g. wrapping). The *context*, **C** , consists of a set of semantic and pragmatic constraints limiting the application of the construction. It can be viewed as a set of *preconditions* that must be satisfied in order for a construction to be used in parsing. The *message*, **M** , describes the meaning of the construction, via a set of syntactic, semantic and pragmatic constraints.

To make this concrete, let us consider a few examples. We begin with a simple "command construction" consisting of an action verb followed by its argument.

$$\left[\begin{array}{l} \mathbf{N}\ :\ sent(cmnd, v.np) \\ \left[\begin{array}{lll} \mathbf{C} & : & [< hr\ attends >= sr] \\ \mathbf{V} & : & \left[\begin{array}{lcl} struc & = & (V.NP) \\ < V\ cons\_n > & = & verb \\ < V\ \mathbf{M}\ \ v\_type > & = & action\_verb \\ < NP\ cons\_n > & = & np \end{array}\right] \\ \mathbf{M} & : & \left[\begin{array}{lcl} sem\_cat & = & command \\ action\_type & = & < V\ \mathbf{M}\ \ sem\_type > \\ action\_object & = & < NP\ \mathbf{M}\ \ sem\_type > \\ agent & = & hr \end{array}\right] \end{array}\right] \end{array}\right]$$

The context of the construction describes all situations in which the the hearer *hr* (human or machine) is paying attention to the speaker *sr* (a "ready" state). The feature *struc* is a list of variables and/or words/tokens; it is used to to describe the structure of a construction, and its role is similar to a production in a generative grammar. (We will write names of variables in capital letters, e.g. NP, inside matrices of constructions). The attribute *cons_n* gives the name of a construction that could be assigned to a string. We use it here to say that the form of the construction can be described as a concatenation of two strings, of which one is a verb (construction) and the other an np (construction). Furthermore, the verb type $< V\ \mathbf{M}\ \ v\_type >$ is "action_verb". (The expression $< V\ \mathbf{M}\ \ v\_type >$ should be read "the *v_type* of the message of V").

The message **M** describes the meaning of the construction as that of a command in which the type of action is described by the meaning of the verb, and the object of the action is given by the meaning of the noun phrase. The attribute *sem_type* stands for the "semantic type" and we identify it currently with the word sense. Thus "erase the file" is understood as a command to *delete* the file, if $< erase\ \mathbf{M}\ \ sem\_type >= delete$, but "erase the picture" might refer to the type of action associated with *rub_out*. In both cases the hearer *hr* is supposed to be the agent of the action. (For commands with pronouns such as: "send it", the meaning of "it" would have to be computed from previous discourse context and/or background knowledge).

38.0.3.1 The Algorithm

The algorithm is quite simple. At each step the following routine is called

```
Parse one sentence:
 1. Read sentence/string.
 2. Parse sentence using
    (a) constructions
    (b) background knowledge
    (c) contextual parameters
 3. Compute attributes important for application
 4. Update current context
```

It is implemented using a chart parser which consults with background knowledge before adding new active edges. The semantic interpreter (point 3) uses the knowledge of the application and current context (e.g. what was the question about) to interpret the string (e.g. an answer that is a fragment with otherwise multiple interpretations). In point 4, contextual variables are updated (e.g. that the context does not contain a question, or that the current action is moving a meeting).

Obviously, a crucial thing in the working of dialog system is *dialog management*, i.e. being able to:
(a) take an order, and figure it out ("set up an appointment");
(b) deal with parameters of the order ("in the afternoon");
(c) ask for parameters ("is it important?");
(d) deal with a change in the parameters ("make it 6");
(e) deal with a change in the order ("no, show my email");
(f) answer simple questions ("do I have any meetings today?");
(g) recover from speech recognition errors ("at what time, two or eight?").
Steps (1)-(4) are part of such a dialog component, which although only partially implemented at this point, shows examples of all possible conversation turns discussed e.g. in Winograd and Flores 1986 p.64ff (`request, promise, counter, declare, assert, reject, ...`). The topic of dialog management will be discussed elsewhere; at this point it is mentioned only to point out some of the contextual factors that help in computing the meanings of utterances; that is, the place in the dialog structure helps in determining the meaning of an utterance.

38.1 Language as a Set of Constructions

In typical linguistic approaches to semantics, a syntactic tree is compositionally mapped into a semantic representation, typically restricted to predicate-argument structure, and then the semantic structure is interpreted within a pragmatic context. In computational linguistics a convincing argument has been advanced that by intertwining syntax and semantics we should get fewer and better parses faster (Lytinen 1991).

A more radical approach to the interaction of syntax and semantics has recently been proposed in Zadrozny and Manaster-Ramer 1994 and Jurafsky 1992. Language is modeled there as a set of constructions (cf. also Fillmore et al. 1988 and Bloomfield 1933). In that model there is no separate syntax, since constructions encode both form and meaning. We assume that each word sense requires a separate semantic description; the same is true about each idiom, open idiom (Fillmore et al. 1988), and a phrasal construction. This means that any grammar of of constructions is non-lexicalist.

Each construction explicitly defines the meaning function taking the meanings of its subconstructions as arguments. That is, we make each semantic function as complicated as linguistically necessary, but their mode of combination is restricted. In the construction-based model semantics is "compositional" and "systematic", but there is no homomorphism from syntax to semantics, because there is no syntax to speak of. It is "compositional", because the meaning of a construction is a function of the meanings of its parts and their mode of combination. (Note that such a function is different for different constructions, and each construction defines its own mode of combination). And it is systematic, because the modes of combination are not arbitrary, as they have to be linguistically justified.

Consider (slightly edited versions of) the messages of some constructions that can be handled by the system (where the arrow is used to represent the structure, *struc*, of constructions):

```
pp(Prep, place) -> [prep(X), np(place(_))],
   msg([[den ,  [location, NPm ]],
                [position, X ] ]),
   veh([...], [ X=in | X=at  ])

pp(in, state(emotion)) ->
                   [prep(in), bare_np(state(emotion))],
   msg([[den , [state(emotion), NPm]] ])
```

In both cases we have a PP construction, where its main meaning `[den,_]` is explicitly specified in terms of the meanings of its parts; in the first example the variable `NPm` is bound to the meaning of `np(place(_))`, and because of the presence of "in" or "at" its value describes *location*. The second example is supposed to account for the construction describing emotional states such as *in love*, and the variable `NPm` is bound to the meaning of *bare_np(state(emotion))* construction.

The next construction shows how subconstructions can be combined into a construction that describes (one type of) a sentence consisting of subject-verb-complement. (E.g. as in *A person is at a table*).

```
s(assert(svc)) -> [np(_), v([be]), pp(X,place)],
msg([ [ den, [situation ,
                  [agent , NPm], [location, PPm]]
veh([...], [ agreement(NPv, Vv, [number, person]])
```

I this case, the result is a *situation* (cf. Barwise and Etchemendy 1987, Devlin 1991). The subconstructions can be combined if some constraints on forms (vehicles) are satisfied, e.g. the agreement. The temporal attribute of the situation is a function of the tense of the verb, and should appear in the description, however it is not necessary to make the point about systematicity. Also notice that a similar construction

```
s(assert(svc)) -> [np(_), v([be]), pp(at,time)]
```

(*The meeting is at 5 pm*) gives explicitly a temporal attribute of a situation.

One could ask why have two or more separate constructions instead of one construction of the form

```
s(assert(svc)) -> [np(_), v([be]), pp(at,X)]
```

with a `case X ...` statement in its message. The answer is that we have adopted a rule that if one structure has two meanings, one should try to divide it into two constructions. In this case this is done by looking at the types of PPs. For arguments against disjunctive rules we refer the reader to Manaster-Ramer and Zadrozny 1990.

38.2 Advantages of Small Domains

In this section we discuss the advantages of small discourse domains. In Zadrozny 1994a we introduce some measures for discourse complexity based on Kolmogorov complexity (cf. Li and P.M.B.Vitanyi 1992), but intuitively we know which domains are small. The purpose of this section is to show how their small size can be explicitly exploited.

38.2.1 Only a Small Percent of Constructions with each Preposition is Needed

It turns out that for a limited domain only a small percent of constructions with each preposition is needed. For instance, for the task of scheduling a room we need 5 out of 30 constructions with "for" mentioned in Collins-Cobuild dictionary (Sinclair 1987). And note that among all prepositions the class of meanings that can be expressed using "for" is one of the least restricted. More specifically, we have to be able to parse the following list of constructions (where the last of the constructions is not mentioned in the dictionary):

something for someone — user Y of X —

"cancel the room for Martin"
do X for someone — beneficiary — "reserve a room for Maria"
(Maria was supposed to do it)
something for something — purpose — "a room for a meeting"
for Time(hour/min/days) — duration — "for 2 hours"
for Date — date — "for August 13"
a room for X people — capacity

From various versions of "from", for the room scheduler we need altogether 3 out 26 constructions mentioned in Collins-Cobuild:

from X to/till/until Y — duration/ begin X; end Y
from Time/Date — beginning X
from 17 to 20 people — range

For "with" this fraction is 3/30, and for "at" 2/24; and the difference between the various usages is in most cases easily computable from the subcategorization patterns and the discourse context (cf. Section 5.2 for a more difficult case).

This observation is not limited to prepositional phrases. The same pattern holds for constructions with the verb "to be", "to have", and many phrasal constructions. But notice that while the domain selects constructions which makes sense there, the constructions do not explicitly mention the domain. Thus they are reusable; they encode general linguistic knowledge.

38.2.2 Ontology: Limited Number of Semantic Categories = Bounds on Background Knowledge

The second advantage of a limited domain lies in the relatively small number of semantic categories. For example, for the domain of calendars the number of concepts is less than 100; for room scheduling it is about 20. Even for a relatively complex office application, say, WordPerfect Office 4.0, the number of semantic categories is between 200 and 500 (the number depends what counts as a category, and what is merely a feature).

Why this is important? Because not only do we need a set of semantic categories, but also we have to encode background knowledge about them. For instance, given the concept of "range" with its "beginning", "end" and "measure"

$$[range, [[[beginning, \_], [end, \_]], [measure, \_]]]]$$

we should have a constraint that the value of "beginning" is smaller than the value of "end". We should know that two different meetings cannot occupy the same room in overlapping periods of time, we should know the number of days in any month, and that meetings are typically scheduled after the current date, etc.

One should ask how many such facts we need. There is evidence (Graesser 1981, Crothers 1979, Zadrozny and Jensen 1991) that the ratio of the number of words to the number of facts necessary to understand sentences with them is about 10:1. In the absence of large bodies of computer accessible common-sense knowledge, this makes the enterprise of building NLU systems for small domains feasible. Thus the advantage of limited domains lies in the fact that background knowledge about them can be organized, hand-coded and tested.

38.2.3 Ontology: Semantic Categories = Parameters of Situations

Although the number of concepts we use is limited, our ontology is richer than the ones discussed in situation semantics literature (e.g. Devlin 1991, Fenstad et al. 1987, Barwise and Etchemendy 1987). Of more specific types that are not dealt with in situation semantics we need for example all calendar concepts (`month, day, ...`) such concepts as: `user`, `capacity`, `measure`, `duration`, `range`, `beginning`, `end`, etc. Also, there are concepts such as `state(emotion)` that are not needed for our particular applications, but seem necessary to account for some PP constructions.

We believe that any semantic category can be a parameter of a situation. "Technical" categories such as `date` are easier to represent than "psychological" or "social" categories. However, in a limited domain even the concepts which refer to a social function, e.g. `participant(s)` have their normal meanings severely restricted, and their parts that are relevant to the set of tasks that a program should handle can be axiomatized.

With respect to the theories of meaning, while we share many of the intuitions expressed by Lakoff 1987, and we agree with many of his criticism of the traditional model theoretic semantics, including some of his criticism of situation semantics, we believe that for limited domains we do not have better tools at present.

38.3 Computing Single Meanings

We will now illustrate the working of the semantic interpreter and the parser, and the handling of ambiguous expressions.

38.3.1 Example Sentences that can be Handled by the System

```
| ?- sem.
|: i want to set up an appointment with leora at 5 pm on the
   14th of march in the ibm cafeteria.
```

```
***Slots:
[  [  action_name schedule]
   [  event_name
      [  an appointment]
   [  event_time
      [  [  minute 0] [ hour 17]
         [  day 14]   [ month 3]
   [  event_place
      [  in the ibm cafeteria]
   [  event_partner
      [  leora]
```

In addition to sentences (and their fragments) dealing with times, places and appointments, the system can also handle a range of other constructions with PPs, including idioms such as *John is at lunch.*

The resulting attribute-value structure that we see above is obtained from the result of parsing. The parser computes the meaning of a sentence, but it does not assign values to slots, because types of slots are application dependent. We will illustrate the working of the parser with a shorter example.

```
| ?- sem.
|: schedule a lunch at 5 in my office.
...
 ---filtered out [np(time(hour)),pp_list(in,place)] --
...
```

Here, the possibility of "place" modifying "time", as in *In winter, 4:30 pm in Kiruna might be depressing*, is excluded. We now show some of the information that is available in the parse chart.

```
| ?- li.
  Chart results: INACTIVE EDGES
* 1,3,[a,lunch] :
  np(event(meal)) -> [det(a_an),n(event(meal))]
  [[type,event(meal)],[den,lunch],
   [mods,[[det,a]]], [quantity,1]]
```

As we can see the semantic representation is pretty straightforward; it consists of the type (i.e. parameter) and its value, plus some additional information which is, or might be, relevant.

In the next few lines we can see that parses might be ambiguous. However, if from the point of view of the working of the semantic engine, we obtain the same slots with the same values, such ambiguities do not have to be filtered out.

```
* 3,8,[at,5,in,my,office] :
  pp_list(in,place) -> [pp(at,time(hour)),pp_list(in,place)]
  [[pp_msg,[[type,place],[den,office],[position,in]]],
   [pp_msg,[[ [type,time(hour)],
              [den,[[hour,[5,am_or_pm]],[minute,0]]]],
              [position(time),at]]]
* 3,8,[at,5,in,my,office] :
  pp_list(in,place) -> [pp(at,time(hour)),pp_list(in,place)]
  [[pp_msg,[[type,place],[den,office],[position,in]]],
   [pp_msg,[[type,time(hour)],
            [den,[[hour,[5,am_or_pm]],[minute,0]]],
            [position(time),at]]]]
* 3,8,[at,5,in,my,office] :
  pp_list(at,time(hour)) ->
                [pp_list(at,time(hour)),pp(in,place)]
  [[pp_msg,[[[type,time(hour)],
             [den,[[hour,[5,am_or_pm]],[minute,0]]]],
             [position(time),at]],
   [pp_msg,[[type,place],[den,office],[position,in]]]]
* 1,8,[a,lunch,at,5,in,my,office] :
  np_list(event) -> [np(event),pp_list(at,time(hour))]
```

Above, the semantics of the phrase would be a concatenation of the respective lists of values, with the message of the `pp_list` becoming part of the `mods` of the `np_list`. Below, the `np_list` is an object of the vp, which in turn is intepreted as an imperative sentence.

```
* 0,8,[schedule,a,lunch,at,5,in,my,office] :
  vp([]) -> [v([]),np_list(event)]
* 0,8,[schedule,a,lunch,at,5,in,my,office] :
  s(imp([])) -> [vp([])]
```

38.3.2 The Ambiguities of "from 10 to 12"

How do we distinguish between the different meanings of the phrase *from 10 to 12*? The context of previous utterance or a task allows us to distinguish between `time`, `capacity`, `office number`. That is, the context can distinguish between the first two readings, if the phrase is an answer to a question; the type of the question determines the reading of the answer. However some ambiguity might remain, e.g. *move the meeting from 10 to 12* might refer to time or room number; moreover the time reading can be ambiguous between the obvious reading and the reading *the meeting (that begins at) 11:50.* The last reading can be eliminated either by forbidding postmodification of nouns by temporal expressions, or by checking what meetings have been scheduled, or by some preference metrics.

Sometimes the knowledge of the application helps. The time expressions *4 to 12* must mean "beginning at 4 and ending at 12" than "11:56" in the context of a question about the time of a meeting. We can completely eliminate the second reading, because all meetings in *xdiary* are scheduled in 5 minute increments.

38.4 Comparisons with Related Work on PPs

We have shown how to deal with the problem of computing the meanings of sentences with PPs. We owe our success to the methodology we described above. In the next few paragraphs we make a short comparison with two pieces of relevant work.

38.4.1 Colban

Colban 1987 was probably the first person to explicitly discuss the problem of meaning of PPs in situation semantics. We believe we have improved on his work in the following ways:

1. We have developed a method for encoding PPs as constructions, (using domain ontologies in that process), and for parsing utterances with PPs.
2. Our system handles about 50 types of PPs (half of them needed for scheduling tasks, the other half to experiment with), and 10 idioms.
3. It handles successfully their combinations (in the domain of calendars); hence it is able to deal with ambiguities.
4. It is implemented and it works with real programs.

However Colban raises an interesting point that we have not explicitly discussed yet, namely, the difference between PPs that describe arguments of verbs, and those that describe adjuncts. This difference is implicitly in the grammar of constructions, where subcategorization patterns for verbs (or phrasal constructions) describe the types and values of arguments of verbs as a function of the messages of subconstructions.

38.4.2 Jensen and Binot

Jensen and Binot 1988 suggest that on-line dictionaries should be used as a source of background knowledge when dealing with the PP–attachment problem. This is true in many instances, although in the case of calendars and other more technical domains, background knowledge is not in the dictionaries.

To make a short comparison between their work and ours, we first note that since there is no separate syntactic component in our grammar, we do not face the PP–attachment problem. However, we have the problem of multiple semantic interpretations. The two are related but not the same, e.g. the meaning of a PP is not only a function of its position, and sometimes the semantic analysis is the same no matter where a PP has been attached (Section 5.1).

Second, the two ideas, i.e. using on-line sources of background knowledge and using a dictionary of constructions, complement each other. The same is true for deriving this background knowledge, manually vs. automatically. That is, being able to process the contents of on-line dictionaries would help a lot in building dictionaries of constructions, and our construction formalism can be one of the tools for representing knowledge contained there. Furthermore, in any concrete application some amount of manual labor is necessary to encode the relevant background knowledge.

Third, using an ontology in linguistic descriptions can be seen as a refinement of the idea of using background knowledge to interpret results of syntactic parsing. That is, using a syntactic grammar does not preclude using information about the ontological status of discourse entities in the later stage of processing.

38.5 Summary

Computing semantics of sentences (utterances) in small domains is not a trivial task, because it involves both deciding that the parameters are relevant, and computing their values in the process of finding the meaning of an utterance. We have illustrated both aspects by examples, in particular, by showing how to compute the meanings of sentences with prepositional phrases. Our method uses the representation of linguistic knowledge as a grammar of construction, i.e. data structures encodes the relevant aspect of form, meaning and context. It takes advantage of the possibility of using domain ontologies as types of parameters of situations.

We have argued that the method works, because (a) The number of constructions, including those involving PPs, is limited for any "small" domain, hence they can be explicitly encoded and tested; (b) Background knowledge needed for understanding them can be encoded, too; in particular, constraints on interactions of PPs can be explicitly stated and tested.

Summarizing, this work makes two contributions to situation theory. The first one is the observation that any element of the ontology of a domain can be a parameter of a situation. The second one lies in showing how to compute the values of those parameters of situations from NL utterances.

References

Barwise, Jon, and John Etchemendy. 1987. *The Liar*. New York, NY: Oxford University Press.

Bloomfield, L. 1933. *Language*. Chicago, IL: The University of Chicago Press.

Colban, E. 1987. Prepositional Phrases in Situation Schemata. In *Situations, Language and Logic*, ed. J.E. Fenstad, P.-K. Halvorsen, T. Langholm, and J. van Benthem. 133–156. Dordrecht, Holland: D. Reidel.

Crothers, E.J. 1979. *Paragraph Structure Inference.* Norwood, New Jersey: Ablex Publishing Corp.

Devlin, K. 1991. *Logic and Information.* Cambridge: Cambridge University Press.

Fenstad, J.E., P.-K. Halvorsen, T. Langholm, and J. van Benthem. 1987. *Situations, Language and Logic.* Dordrecht, Holland: D. Reidel.

Fillmore, Ch.J., P. Kay, and M.C. O'Connor. 1988. Regularity and idiomaticity in grammatical constructions. *Language* 64(3):501–538.

Gazdar, G., and Ch. Mellish. 1989. *Natural Language Processing in POP-11/Prolog/Lisp.* Wokingham, England: Addison-Wesley.

Graesser, A.C. 1981. *Prose Comprehension Beyond the Word.* New York, NY: Springer.

Jensen, K., and J.-L. Binot. 1988. Disambiguating Prepositional Phrase Attachments by Using On-line Dictionary Definitions. *Computational Linguistics* 13(3-4):251–260.

Jurafsky, D. 1992. *An On-line Computational Model of Sentence Interpretation.* Doctoral dissertation, University of California, Berkeley. Report No. UCB/CSD 92/676.

Lakoff, G. 1987. *Women, Fire and Dangerous Things.* Chicago, Illinois: The University of Chicago Press.

Li, M., and P.M.B.Vitanyi. 1992. Inductive Reasoning and Kolmogorov Complexity. *Journal of Commputer and System Sciences* 44(2):343–384.

Lytinen, S.L. 1991. Semantics-first Natural Language Processing. In *Proceedings AAAI-91*, 111–116. Anaheim, CA.

Manaster-Ramer, A., and W. Zadrozny. 1990. Expressive Power of Grammatical Formalisms. In *Proceedings of Coling-90*, 195–200. Helsinki, Finnland. Universitas Helsingiensis.

Sinclair, J. (ed.). 1987. *Collins-Cobuild English Language Dictionary.* London: Collins ELT.

Winograd, T., and F. Flores. 1986. *Understanding Computers and Cognition.* Norwood, NJ: Ablex.

Zadrozny, W. 1994a. Parsing with constructions and ontology. In *Proc. AAAI Fall Symposium on Knowledge Representation for Natural Language Processing in Implemented Systems.* New Orleans, Louisiana: AAAI.

Zadrozny, W. 1994b. Reasoning with background knowledge – a three-level theory. *Computational Intelligence* 10(2):150–184.

Zadrozny, W., and K. Jensen. 1991. Semantics of Paragraphs. *Computational Linguistics* 17(2):171–210.

Zadrozny, W., and A. Manaster-Ramer. 1994. The Significance of Constructions. IBM Research Division Technical Report RC 20002.

Zadrozny, W., M. Szummer, S. Jarecki, D.E. Johnson, and L. Morgenstern. 1994. NL understanding with a grammar of constructions. *Proc. Coling'94*.

Index

Name Index