ROUTLEDGE STUDIES IN AMERICAN PHILOSOPHY

Sellars and Contemporary Philosophy

Edited by
David Pereplyotchik
and Deborah R. Barnbaum

Sellars and Contemporary Philosophy

Wilfrid Sellars made profound and lasting contributions to nearly every area of philosophy. The aim of this collection is to highlight the continuing importance of Sellars' work to contemporary debates. The contributors include several luminaries in Sellars scholarship, as well as members of the new generation whose work demonstrates the lasting power of Sellars' ideas. Papers by O'Shea and Koons develop Sellars' underexplored views concerning ethics, practical reasoning, and free will, with an emphasis on his longstanding engagement with Kant. Sachs, Hicks, and Pereplyotchik relate Sellars' views of mental phenomena to current topics in cognitive science and philosophy of mind. Fink, deVries, Price, Macbeth, Christias, and Brandom grapple with traditional Sellarsian themes, including meaning, truth, existence, and objectivity. Brandhoff provides an original account of the evolution of Sellars' philosophy of language and his project of "pure pragmatics." The volume concludes with an author-meets-critics section centered around Robert Brandom's recent book, *From Empiricism to Expressivism: Brandom Reads Sellars*, with original commentaries and replies.

David Pereplyotchik is assistant professor of philosophy at Kent State University. He received his PhD in philosophy, with a concentration in cognitive science, from the CUNY Graduate Center. He is the author of *Psychosyntax: The Nature of Grammar and Its Place in the Mind* (2017) and is an active member of the Wilfrid Sellars Society.

Deborah R. Barnbaum is professor and chair of philosophy at Kent State University. She received her PhD in philosophy from the University of Massachusetts in 1996. She is the author of *The Ethics of Autism* (2008) and the co-author of *Research Ethics: Text and Readings*, with Michael Byron (2001).

Routledge Studies in American Philosophy

Edited by Willem deVries
University of New Hampshire, USA
and
Henry Jackman
York University, Canada

Sellars and Contemporary Philosophy

Edited by David Pereplyotchik
and Deborah R. Barnbaum

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2017
by Routledge
711 Third Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging in Publication Data
Names: Pereplyotchik, David, editor. | Barnbaum, Deborah R., 1967– editor.
Title: Sellars and contemporary philosophy / edited by David Pereplyotchik and Deborah R. Barnbaum.
Description: New York : Routledge, 2016. | Series: Routledge studies in American philosophy; 7 | Includes bibliographical references and index.
Identifiers: LCCN 2016036038 | ISBN 9781138670624 (hardback : alk. paper)
Subjects: LCSH: Sellars, Wilfrid.
Classification: LCC B945.S444 S46 2016 | DDC 191—dc23
LC record available at https://lccn.loc.gov/2016036038

ISBN: 978-1-138-67062-4 (hbk)
ISBN: 978-1-315-61754-1 (ebk)

Typeset in Sabon
by Apex CoVantage, LLC

Contents

Acknowledgments

We have many people to thank for making this book possible. We owe our greatest appreciation to Willem deVries, who, on March 16, 2014, agreed to come to Kent State University for what at the time was merely the vaguest idea of a conference on Wilfrid Sellars. One year later, an extraordinary group of scholars assembled for the April 30 through May 2, 2015, *Sellars in a New Generation* conference. Bill's contributions were invaluable at every step: agreeing to serve as one of the keynote speakers at the conference; helping to brainstorm the theme of the conference; reviewing and assisting in the selection of papers; participating in the conference with his signature combination of wit, aplomb, and enthusiasm; and finally, offering feedback on the book proposal. Without Bill's support we would not have been able to put together this volume. Our gratitude is immeasurable and heartfelt.

In addition to Bill deVries, we wish to also acknowledge the work of Deborah C. Smith and Gene Pendleton, both of Kent State University, in helping us to choose the papers presented at the conference. They generously gave of their time, and philosophical acumen, to find the excellent papers collected herein.

Alexander Haas compiled the index for this volume. We thank him for his outstanding attention to detail.

The editorial team at Routledge has been very helpful. In particular, Andrew Weckenmann has been with us through every stage of the process. His counsel made compiling this volume all the more enjoyable. We are also grateful to Harvard University Press for granting permission to reprint portions of the first chapter of Robert B. Brandom's From *Empiricism to Expressivism: Brandom Reads Sellars*.

Kent State University's significant financial contributions made this volume possible. James L. Blank, professor and dean of the College of Arts and Sciences, generously funded the *Sellars in a New Generation* conference. The Office of Research and Graduate Studies provided financial support to Alexander Haas. We are also grateful to the entire faculty of the Kent State University Philosophy Department, who participated in the conference. We are privileged to call such a supportive environment, populated by engaged colleagues, our home.

We wish to thank each of the contributors for their participation in the conference and in the compilation of this volume. It was our pleasure and honor to work with each of them. One goal in reproducing the Author Meets Critics section was to capture, for readers, the vibrant engagement that resulted from assembling a room full of veteran and emerging Sellars scholars. The audience at the conference—a few of whom are part of the Author Meets Critics discussion—were invaluable partners in this enterprise. Sellars scholarship is only as good as Sellars scholars themselves—a cadre of intelligent, impassioned, humorous, and charming philosophers. Bringing them together was an honor.

Finally, on both a professional and personal level we wish to thank our respective spouses, Aoife Mooney and Gene Pendleton. As mentioned above, Gene Pendleton contributed to the selection of the papers for the conference. Aoife Mooney designed the conference programs and the promotional materials. In addition to taking advantage of their copious professional gifts, we are fortunate to have their support in our work.

David Pereplyotchik
Deborah R. Barnbaum
Kent State University

A Note on References

All references to the works of Wilfrid Sellars in this volume are abbreviated using the standard conventions found at plato.standford.edu/entries/Sellars/ (accessed June 29, 2016). Full citations may be found in the reference sections of each paper.

Books

EPH:	*Essays in Philosophy and Its History*
ISR:	*In the Space of Reasons: Selected Essays of Wilfrid Sellars*
KTM:	*Kant's Transcendental Metaphysics: Sellars' Cassirer Lectures Notes and Other Essays*
SM:	*Science and Metaphysics: Variations on Kantian Themes*
NAO:	*Naturalism and Ontology*
PPME:	*Philosophical Perspectives: Metaphysics and Epistemology*
SPR:	*Science, Perception, and Reality*
WSNDL:	*The Notre Dame Lectures*

Essays

BD:	"Berkeley and Descartes: Reflections on the 'New Way of Ideas'"
BLM:	"Behaviorism, Language and Meaning"
CC:	"Conceptual Change"
CDCM:	"Counterfactuals, Dispositions, and the Causal Modalities"
CIL:	"Concepts As Involving Laws And Inconceivable Without Them"
EAE:	"Empiricism and Abstract Entities"
ENWW:	"Epistemology and the New Way of Words"
EPM:	"Empiricism and the Philosophy of Mind"
FCET:	"Form and Content in Ethical Theory"
FMPP:	"Foundations for a metaphysics of pure process: The Carus lectures"
I:	"'. . . this I or he or it (the thing) which thinks . . .'"
ILE:	"The Identity of Linguistic Expressions and the Paradox of Analysis"

IM:	"Inference and Meaning"
ITM:	"Intentionality and the Mental"
ITSA:	"Is there a Synthetic A Priori?"
KBDW:	"On Knowing the Better and Doing the Worse"
KTE:	"Some Remarks on Kant's Theory of Experience"
LRB:	"Language, Rules and Behavior"
LT:	"The Language of Theories"
LTC:	"Language as Thought and as Communication,"
MEV:	"Mental Events"
MFC:	"Meaning as Functional Classification"
MGEC:	"More on Givenness and Explanatory Coherence"
MP:	"Metaphysics and the Concept of a Person"
OAFP:	"On Accepting First Principles"
OMP:	"'Ought' and Moral Principles"
ORAV:	"On Reasoning about Values"
PHM:	"Phenomenalism"
PPE:	"Pure Pragmatics and Epistemology"
PRE:	"Presupposing"
PSIM:	"Philosophy and the Scientific Image of Man"
QMSP:	"Quotation Marks, Sentences and Propositions"
RNWW:	"Realism and the New Way of Words"
SE:	"Science and Ethics"
SK:	"The Structure of Knowledge"
SRI:	"Scientific Realism or Irenic Instrumentalism: A Critique of Nagel and Feyerabend on Theoretical Explanation"
SRPC:	"Some Reflections on Perceptual Consciousness"
SSIS:	"Science, Sense Impressions, And Sensa: A Reply To Cornman"
TC:	"Truth and 'Correspondence'"
TE:	"Theoretical Explanation"

Figure 0.1 The above photo was taken at the Sellars in a New Generation Conference on May 2, 2015, at Kent State University. Pictured from left to right (Top row): Jeremy Randel Koons, David Pereplyotchik, Boris Brandhoff, Dionysis Christias, Huw Price, Danielle Macbeth, Carl B. Sachs, Willem A. deVries. (Bottom row): Robert B. Brandom, Michael R. Hicks, James O'Shea, David Rosenthal. (Kevin Fink not pictured.)

Introduction

David Pereplyotchik and
Deborah R. Barnbaum

The work of Wilfrid Sellars touched nearly every area of philosophy, from traditional debates in metaphysics, epistemology, and ethics, to contemporary issues in philosophy of language and mind. Sellars' contributions to each of these fields were strikingly original, historically informed, and deeply systematic. His influence on the trajectory of American philosophy was profound.

In recent years, Sellars scholarship has enjoyed a renewed interest. A formal society dedicated to the study of his work now meets regularly, and many conferences have been devoted to exploring and extending his work (deVries, 2009; O'Shea, 2016). Three recent books lay out Sellars' main lines of thought (deVries, 2005; O'Shea, 2007; Rosenberg, 2007), another offers a critical take (Brandom, 2015), and many others draw heavily on Sellars in deepening our understanding of consciousness (Rosenthal, 2005), perception (Coates, 2007), meaning (Brandom, 2008), ontology (Price, 2011), and the history of philosophy (Brandom, 2002; O'Shea, 2014; Olen, 2016).

In May 2015, the Philosophy Department at Kent State University held the *Sellars in a New Generation* conference. The aim was to promote Sellars scholarship by highlighting the relevance of Sellars' work to the questions that philosophy faces today. The papers collected in this volume were, with one exception, presented at the conference, though they have been edited to reflect the outcomes of the discussions and debates that took place there.

In an effort to bring young researchers into the community of Sellars scholars, we began the conference with a graduate workshop, where early-career philosophers had an opportunity to present their work on Sellars in the company of more established colleagues. Two of the graduate student submissions, by Boris Brandhoff and Kevin Fink, have been selected for publication in this volume. Their outstanding discussions—one critical of Sellars' views on color and one sympathetic to his project of "pure pragmatics"—are a model of the Sellars scholarship that we hope to see in the new generation.

The conference concluded with an author-meets-critics session centered on Robert B. Brandom's recent book, *From Empiricism to Expressivism: Brandom Reads Sellars*. The portion of the book that Brandom presented at

the conference is reproduced here. In it, he develops his critical interpretation of Sellars by clarifying and extending one of Sellars' core claims, and then using it to argue against another. If he is right, then the major pillars of Sellars' thought are in deep tension with one another. Brandom's arguments are the topic of a paper by Dionysis Christias, as well as two commentaries by Willem A. deVries and James O'Shea, which appear in this volume alongside Brandom's replies and audience discussion.

In selecting the papers for this collection, our goal was to diversify the range of topics addressed by the contributors, so as to illustrate and extend the impressive scope of Sellars' work. Though this strategy proved successful, the resulting diversity presents a difficulty of finding useful categories into which to group the papers. This is exacerbated by the fact that many of the contributors adapted for their purposes a small batch of core Sellarsian themes, including naturalism, expressivism, pragmatism, functionalism, realism, and a Kantian approach to moral and metaphysical issues. We have chosen a categorization scheme that we hope is both useful and illuminating in the face of this mix of diversity and overlap. The papers are here presented as falling into three sections: (1) Ethics, Moral Reasoning, and Free Will; (2) Philosophy of Language and Mind; and (3) Metaphysics and Epistemology. The remainder of this introduction situates and summarizes each of the papers and points to a number of recurring themes.

The papers in the first section, by James O'Shea and Jeremy Randel Koons, are contributions to the underexplored area of Sellars' moral philosophy—including his views on ethics, metaethics, moral agency, and practical reasoning.

The fruits of Sellars' deep and longstanding engagement with Kant come to the foreground in James O'Shea's paper, "Thought, Freedom, and Embodiment in Kant and Sellars." O'Shea's discussion provides a close examination of the ideas that Sellars set out in his APA Presidential Address (Sellars, 1972), in which he offered an interpretation of Kant's views on free will and moral reasoning, and then propounded his own Kantian naturalism. Noting that some commentators have misinterpreted Sellars as implausibly imputing materialist views to Kant, O'Shea argues that Sellars was in fact offering a *reconstructive* defense of Kant's views, by identifying adjustments to them that could render them consistent with a materialist metaphysics. He points out also that Sellars adopted an interpretive strategy according to which Kant's epistemology in the first *Critique* should be viewed as provisionally abstracting away from various empirical matters (e.g., embodiment) that are later reintroduced and treated more fully in the third *Critique*. O'Shea argues that Sellars' outlook is "crucial for correctly interpreting Kant's critical philosophy across the board, and in particular for questions concerning how to understand Kant's *applications* of his transcendental principles in his writings beyond the three *Critiques*." He goes on to illustrate how Sellars' account of practical reasoning allows for a reconciliation between Kantian and consequentialist views, by incorporating the idea that the most general

maxim on which the virtuous ground their moral and practical reasoning is the maximization of the welfare of all rational creatures. Finally, O'Shea explains how Sellars' naturalistic account of practical reasoning can address Kant's deepest concerns about the incompatibility of free will and mechanistic determinism.

Like O'Shea, Jeremy Randel Koons brings out several Kantian elements of Sellars' moral philosophy in his paper "Toward a Sellarsian Ethics for the 21st Century." These include the idea that reason has motivational force independently of desires and passions, and that "moral principles primarily concern the consequences of anybody acting in a certain way in a certain kind of circumstances" (Sellars, 1967: 207). Koons begins his paper by focusing on Sellars' account of practical reasoning, which relies crucially on the notion of a collective intention—what Sellars called a "we-intention." He suggests that this notion can be useful in explicating a species of reasoning that is not individual but rather, collective and cooperative. A proper appreciation of cooperative reasoning, Koons argues, would show that it is both irreducible to individual reasoning and essential to a genuinely moral point of view. In highlighting various features of Sellars' conception of practical reasoning, Koons points out that practical reasoning consists of material, rather than formal, inferences, contrary to what some of Sellars' working examples may suggest. He goes on to explore Sellars' view that some categorical imperatives are *intrinsically* reasonable, despite their having been derived from other imperatives via an inferential process of means-ends reasoning. Drawing on Sellars, McDowell, and Brandom, Koons ends by considering the question of whether there are any intrinsically reasonable categorical imperatives, and the related question of whether there are conclusive grounds for adopting a moral point of view.

The second section of this volume consists of four papers, each devoted to an aspect of Sellars' philosophy of language and philosophy of mind. The papers by Boris Brandhoff and Michael Hicks offer exegetical analyses of Sellars' views on linguistic and mental phenomena—one sympathetic, the other critical. The papers by David Pereplyotchik and Carl B. Sachs both aim to relate Sellars' philosophy of mind to live issues in the philosophy of cognitive science, though in quite different ways.

Brandhoff's paper, "Pure Pragmatics and the Idea of a Metatheoretical Functionalism," advances a contentious exegesis of Sellars' early and middle period, arguing that his core metaphilosophical commitments concerning the project of "pure pragmatics" remained stable in the face of more superficial changes. (For an opposing view, see Olen, 2016.) Brandhoff argues that Sellars' initial characterization of philosophy as a "formal" enterprise was misleading, given what that term would have meant to Sellars' contemporaries, notably Rudolf Carnap. Sellars' concern, Brandhoff points out, is not with language as a set of sign-designs, but rather, with language as set of normative social practices, governed in part by "conformation rules." Correspondingly,

"Sellars' commitment to a formalistic conception of philosophy comes down to a commitment to the idea that the task of philosophy is to clarify, or explicate, the system of functional roles which serves as a norm for our factual linguistic behavior."

In "What Jones Taught the Ryleans," Michael R. Hicks identifies a tension between two core Sellarsian commitments regarding the nature of thought and its relation to public language. On the one hand, Sellars holds that conceptual thought is "essentially public," in the sense that its intentionality must be explicated in terms of the social norms instituted by community practice. On the other hand, Sellars takes thoughts to be covert episodes, which are initially *not* in the public arena, but are eventually *discovered*. Hicks conducts a critical exegesis of Sellars' famous Myth of Jones in *Empiricism and the Philosophy of Mind* and concludes that these two elements of Sellars' position are indeed incompatible: "while Sellars's realism require[s] that conceptual episodes can be overlooked, publicity require[s] that [they] figure in social life." Hicks' diagnosis of the problem touches on Roderick Chisholm's famous objection to Sellars' account of the meaning of public language items (Chisholm and Sellars, 1957). Sellars sought to characterize linguistic meaning without making any reference to covert conceptual episodes, and then to use this account as a model for the theoretical introduction of such episodes. Chisholm objected that this strategy is circular; no account of linguistic meaning can be brought off without at least tacitly appealing to the intentionality of thought, because the latter is conceptually and explanatorily prior to the former. Hicks suggests abandoning Sellars' strategy, but without granting Chisholm's priority thesis. Rather, he argues, we must see public language as *already* a case of conceptual thought. To illustrate his proposal, Hicks points out that understanding, though in some sense a covert process of conceptual thought, is also a recognizably *public* feature of language, at least as important as the overt production of speech acts. If he is right, then we have here a case of conceptual thought that is plausibly regarded as public, in that it is susceptible to social acknowledgment and criticism.

In "Sellars and Psycholinguistics," David Pereplyotchik likewise discusses linguistic comprehension, though from the perspective of cognitive science, which is decidedly more friendly to the notion of covert episodes. Pereplyotchik begins by summarizing the empirical case from contemporary psycholinguistics for positing a range of internal states—what he calls "mental phrase markers"—that causally mediate the process of language comprehension. Drawing on Sellars' flexible and purpose-relative functional-role semantics, he argues that such states can profitably be thought of as a distinctive kind of *representation*. In keeping with Sellars' insights about representational systems, he contends that the representational properties of mental phrase markers are determined in part by the functional roles they play in a creature's internal computational economy,

rather than solely by causal covariation with elements of the environment. Pereplyotchik goes on to make use of several aspects of Sellars' philosophy of mind in articulating the key differences between mental phrase markers and the intentional states that are characteristic of genuine sapience. Unlike personal-level conceptual episodes, the subpersonal states involved in language processing are inexpressible in speech, absent from consciousness, and subject to teleological rather than social norms—features that Pereplyotchik argues are interrelated in theoretically significant ways. In developing a fuller understanding of the distinctive brand of normativity that is characteristic of the subpersonal level, he draws on Daniel Dennett's distinction between intentional-stance and design-stance explanations. Finally, Pereplyotchik highlights Sellars' distinction between ordinary and auxiliary positions in a language game and develops an analogous distinction at the subpersonal level. Mental phrase markers are, on this view, the subpersonal analogues of ordinary positions, while the subpersonal analogues of auxiliary positions are the rules or principles of a hypothetical mental grammar. If the human language-processing system is, in Sellars' terms, a Humean rather than an Aristotelian one, then such auxiliary positions play no role in its operations, and are thus "psychologically real" in only a very attenuated sense.

Carl B. Sachs' paper, "Sentience and Sapience: The Place of Enactive Cognitive Science in Sellarsian Philosophy of Mind," explores the contrast between Sellars' critical realism (developed also by Coates, 2007) and enactivist accounts of perception. Sellars takes sentience and sapience to be "transcendental" structures, which, though posited to explain experience and empirical knowledge, must also be adequately reflected or realized in causal structures. To this end, his critical realist account of the transcendental structure of "sheer receptivity" posits both sense-impressions and the activities of the productive imagination as causally mediating our perceptual encounters with the world. In line with recent embedded/embodied approaches in cognitive science, Sachs suggests that a better account of receptivity would appeal instead to the structural coupling between sensorimotor skills and environmental affordances, a theoretical construct that has its home in enactive accounts of perception. In developing this idea, Sachs addresses a challenge to the effect that sensorimotor skills seem too "active" to ground our receptive faculties, by distinguishing between passivity and receptivity. Our receptive faculties, he proposes, need not be seen as entirely passive. Sachs concludes by exploring the possibility of synthesizing enactivist views of perception with Sellars' critical realism. This hybrid view, which he dubs "embodied critical realism," treats sensorimotor abilities as taking the place of the productive imagination in our account of perceptual experience.

The papers in the third section, by Huw Price, Kevin Fink, Willem deVries, Danielle Macbeth, Dionysis Christias, and Robert Brandom, grapple with

traditional metaphysical and epistemological issues concerning truth, meaning, existence, and objectivity. For Sellars, these issues arise in our attempts to discern the relations between what he called the Manifest and the Scientific images of the world (Sellars, 1963). Sellars used this contrast between the two images as a rubric for unifying the interrelated problems of reconciling minds with bodies, norms with facts, reasons with causes, sensations with brain processes, persons with nature, common sense with science, and appearance with reality. Indeed, Sellars' overall philosophical project can be cast as an attempt to achieve a synoptic vision, in which both the Scientific and the Manifest images are given their due. To do this, Sellars thought, we must come to see how persons can be, on the one hand, natural objects—made up of the same basic entities or processes as everything else in the natural world—but on the other hand, also the subjects of sensory experience and rational thought. Descriptions of persons that appeal to these latter properties raise the so-called "hard problem" of sensory consciousness and the problem of locating intentionality in the causal order. With regard to the latter, Sellars holds that the rationality of persons, and the intentional properties of their states and activities, are both ultimately a matter of how conceptual norms are instituted in social groups. So the problem of intentionality shows up as a special case of the more general problem of reconciling facts and norms—the "is" of the Scientific image and the "ought" of the Manifest image.

Huw Price, in his paper, "Wilfrid Sellars Meets Cambridge Pragmatism," identifies a view common to both Sellars and the pragmatists in the Cambridge tradition, including Ramsey, Wittgenstein, Blackburn, and Price himself. This view, which he calls "Humean expressivism," combines two theses concerning normative, modal, mathematical, and semantic vocabularies. The negative semantic thesis is that the claims made with such vocabularies do not describe or represent the world, and hence are not fact-stating or truth-apt. The positive pragmatic thesis is that such claims are nevertheless subject to rational evaluation, support, assent, and dissent, and are thus "cognitive." Price highlights a problem for this view, namely, the difficulty of drawing a principled distinction between "descriptive" and "cognitive" uses of language. He chronicles Sellars' struggle with this problem, which eventuates in Sellars drawing a sharp distinction between two notions of truth (or perhaps two notions conflated under the one term, 'truth'). The first notion is that of "semantic assertibility" (i.e., correct assertibility in accordance with the semantic rules of a language). This is a generic notion, applying as much to moral, modal, and mathematical claims as to claims about minerals and magnets. The second notion is that of "picturing," which applies more narrowly to only "matter-of-factual" claims. Sellars maintained that these two notions do very different jobs for us and are subject to different criteria of application; picturing is itself a matter-of-factual relation, whereas correct assertibility is not. Price argues that drawing this

distinction allowed Sellars to dissolve the problem facing Humean expressivism by rejecting the negative semantic thesis. If the relevant semantic notion is that of correct assertibility, rather than picturing, then it is simply false that the putatively problematic vocabulary and claims fail to describe, represent, state facts, or be truth-apt. Sellars thus has something to offer the Cambridge pragmatists who are attracted to Humean expressivism. But Price goes on to note that Sellars did not fully appreciate the metaphysically deflationary consequences of adopting the positive pragmatic thesis inherent in that view. He argues that Sellars' distinction ultimately leads to the "global expressivism" and "subject naturalism" that he develops and defends in Price (2011).

Whereas Price's discussion deals with the objectivity of norms and probabilities, Kevin Fink's paper, "An Incoherence in Sellars' Error Theoretical Account of Color Concepts," tackles the objectivity of colors. Fink begins by tracing the stages of Sellars' story about the evolution of our color concepts. In the earliest stages, color is seen as a kind of *stuff* out of which material objects are made, or perhaps as a *part* of those objects. A conceptual revision is then made, so that color comes out instead as a *property* of material objects. In a subsequent stage, the theoretical innovation of the mythical genius Jones (Sellars, 1956) yields a bifurcation between "physical" color, a perceptible property of material objects, and "mental" color, a quality of internal sense-impressions. The final stage of the story represents Sellars' considered view on the matter—namely, that color is a property *only* of sense impressions, *not* of mind-independent material objects. Considerations pertaining to what he called the "ultimate homogeneity" of colors led Sellars to a kind of error theory, according to which our ordinary concept of color is that of a perceptible property, but in fact no such property exists in material objects. Colors are, on this view, "relocated" into the mind. While recognizing that the point about ultimate homogeneity has further troubling consequences for Sellars' metaphysics of mind, Fink ends his exegesis here and begins his critical assessment. He argues that an error-theoretic account of color concepts is actually unavailable to Sellars, for it would be unintelligible in light of Sellars' own theory of conceptual content. Fink points out that, for Sellars, the content of a concept is determined by the nonlogical laws and the "material invariancies" in which it figures. But for Sellars' error theory to be viable, Fink argues, color concepts would have to play a role in substantive, nonlogical laws. The problem is that "mental" colors are only sufficiently involved in such laws when tied to "physical" colors. On their own, any material invariancies in which they might figure would be insufficient to endow them with any content. If this is correct, then Sellars cannot consistently "relocate" color into the mind.

Willem A. deVries approaches the relation between the Scientific and Manifest images in still another way, in his paper, "The Causal Articulation of Practical Reality." His goal in this paper is to determine the bearing

of Sellars' naturalism on the ontological status of the items that constitute what deVries calls "practical reality" (i.e., persons and normatively constituted objects and activities, such as hammers and elections). He argues that genuine agency, as exhibited by, for instance, voting in an election, requires the possession and *unqualified use* of concepts such as "ballot," as well as a first-person conception of oneself as an agent. It follows that no matter how sophisticated a community's scientific understanding becomes, the practical realities of child-rearing, education, and the maintenance of a functioning social order will necessitate the use of concepts from the Manifest image. But does this mean that such concepts refer to things that *really exist*? According to deVries, in discussing such ontological questions, Sellars often appealed—in person, if not in print—to the Eleatic principle that existence is a matter of participating in causal interactions. DeVries goes on to examine Sellars' account of causation and uses it to argue that countenancing the denizens of our practical reality in one's ontology requires giving up on Sellars' "heterodox, super-atomistic, Tractarian process ontology."

Though she does not put it in these terms, Danielle Macbeth's paper, "Natural Truth," can likewise be read as an argument for the existence of a practical reality—that is, for the ontological legitimacy of the Manifest image. Macbeth argues that several core Sellarsian claims jointly entail the existence of what she calls "natural truths." Although not the same for, or available to be grasped by, all rational beings, natural truths are those that are valid for, and available to be grasped by, all rational *human* beings, with our sort of body and form of sensibility. Macbeth points out that the existence of natural truths conflicts with cultural relativism and entails that our powers of perception are powers of knowing, subject to epistemic evaluation. The Sellarsian ideas that Macbeth marshals include the essentially *cultural* nature of language and rationality, the dependence of conceptual meaning on material rules of inference, the applicability of ought-to-do and ought-to-be rules to our inferences and perceptions, and the importance of seeing rational inquiry as a self-correcting *process*, rather than a static structure with foundations. Taken together, these commitments should compel us, Macbeth argues, to construe perception as being rationally accountable to how things *really are*. She argues that the perceptual errors that are revealed by recent studies of implicit bias against women and African-Americans are ones that we are rationally obliged to correct—products of cultural conditioning that we must transcend if we are to see women and African-Americans as they really are. Drawing on these conclusions, Macbeth argues that Sellars is mistaken to claim that the Manifest image delivers mere appearances. For, if the existence of natural truths entails that our powers of perception are powers of knowing, then in perceiving the world through the conceptual framework of the Manifest image, we are accountable to the way the world really is, not merely the way it appears to be.

A common thread running through the papers by Price, Fink, deVries, and Macbeth is that the Manifest image is not happily construed as supplying

us with mere appearances. If these authors are correct, then normative and modal claims are truth-apt (and to that extent factual), colors are real properties of material objects, hammers and elections have a positive ontological status, and perceptual judgments that deploy Manifest-image concepts are rationally accountable to the way that things actually are. This theme is taken up once more by the contributions from Dionysis Christias and Robert Brandom, as well as the accompanying critical commentaries on Brandom's work, by deVries and O'Shea.

In the first chapter of his recent book, *From Empiricism to Expressivism: Brandom Reads Sellars*, Brandom identifies two axial thoughts, around which much of Sellars' philosophy revolves. The first is that Kant's distinction between the phenomenal world of appearances and the noumenal world of things-in-themselves can be usefully lined up with the distinction between the Manifest and Scientific images—the latter serving as the reality that underlies and explains the appearances of the former. Brandom argues forcefully that Sellars' attempt to assimilate this pair of distinctions is a decidedly Bad Idea. And that's because it conflicts, in an unobvious but very important way, with the second of Sellars' two axial thoughts. In Brandom's own words, this is "the Kantian idea that besides concepts whose characteristic expressive job it is to describe and explain empirical goings-on, there are concepts whose characteristic expressive job it is to make explicit necessary structural features of the discursive framework within which alone description and explanation are possible." Brandom is enthusiastic about this idea, and endorses Sellars' contention that such framework-structuring concepts should be understood as metalinguistic in character. Applying the results of his own exploration, in *Between Saying and Doing* (Brandom, 2008), of the systematic relations between meaning and use, Brandom offers an friendly amendment to Sellars' view, by treating the framework-structuring concepts as belonging to a *pragmatic* metalanguage, as opposed to a semantic one.

A particularly striking upshot of this approach is what Brandom calls the "Kant-Sellars thesis" about modality, according to which the use of ordinary descriptive vocabulary presupposes inferential abilities that can be leveraged to permit the legitimate introduction of modal vocabulary. Brandom takes from Sellars the idea that a judgment can only have the kind of intentional content that is characteristic of objective empirical descriptions by being involved in a suitably wide range of material inferences. A sapient creature must have the practical ability to infer such a judgment from some of its collateral commitments, and to refrain from inferring it from others. The capacity to make empirical claims thus rests on an implicit understanding of which claims would remain true in counterfactual circumstances. From one's current commitment to the match being dry, one infers that it will light when struck; but had one believed that there is no oxygen in the room, the same inference would not be drawn. This practical ability, Brandom claims, is all that is needed in order for one to be in a position to learn how to deploy subjunctive conditionals, like "If there were no oxygen in the room,

then the match wouldn't light." For, though such claims do not *describe* a pattern of inference, in the sense of *saying that* one can, will, or must infer the consequent from the antecedent, they do *convey* one's practical commitment the goodness of that inference. That is, they allow one to *express* that commitment verbally—to make explicit in speech what was previously only implicit in practice.

We are now in a position to set out the conflict that Brandom detects between Sellars' two axial thoughts. Brandom points out that the Kant-Sellars thesis about modality has an important bearing on issues concerning the identity relation. If no descriptive empirical concept is "modally insulated," then the identity relation can only hold between items that share *all* of their properties—including modal ones. And since objects that fall under different sortals will rarely have the same modal profiles, it follows that there will be vanishingly few cases of true cross-sortal identity. For instance, no statue can be identical with a lump of clay, no mind can be identical with a brain, no person can be identical with a body, and no table can be identical with a swarm of particles arranged table-wise. But as Brandom notes, one of the most natural ways of interpreting Sellars' view about the ontological primacy of the Scientific image is as a commitment to precisely these sorts of identities. Hence, a line of reasoning that begins with the Kant-Sellars thesis appears to pose serious problems for the idea that the Scientific image can ever *displace* the Manifest image by revealing that the objects of the latter were, all along, identical with objects of the former. In this way, Brandom argues, Sellars' insight about the framework-structuring nature of modal concepts is strictly *incompatible* with his claim that the Manifest image is mere phenomenal appearance, whose underlying noumenal reality consists of things that are literally *identical* with hammers, persons, elections, and the like.

One might of course reply that Sellars' claim about the ontological primacy of the Scientific image need not be construed as a commitment to such cross-sortal *identities*. For instance, Brandom (2014) considers and explicitly argues against another natural interpretation, on which the objects of the Scientific image are seen as the realizers of various functional roles that are specifiable in a Ramsey-sentence consisting entirely of predicates from the Manifest image. An alternative interpretation is developed by Dionysis Christias, in his paper, "Does Brandom's Kant-Sellars Thesis about Modality Undermine Sellars' Scientific Naturalism?" Christias argues that "[t]he descriptive resources of the scientific image are superior to those of the manifest image in matters of ontology because they provide better *explanations* of the world and of our place in it than do those of the scientific image." His proposal is that scientific inquiry proceeds by "*eliminating* manifest-image normative descriptions and explanations" and by changing the modal profiles of the predicates that they contain, thus "changing their very meaning, thereby *recategorizing* them, in their descriptive and explanatory dimension" (emphases in the original). One might object that even if the categorial transition from the Manifest to the Scientific image changes

the sense of these predicates, it might still preserve their reference. Christias respond to this objection by arguing that Sellars rejected relational accounts of both sense and reference, seeing the latter as conceptually dependent on the former. He concludes by providing textual evidence for the claim that Sellars embraced the radical eliminativist consequences of his views, though not without qualification. Eliminativist claims, in Sellars' view, are comments *about* the Manifest image, not *within* it, and they are useful only in the dimension of describing and explaining. They ignore both the practical dimension and the diachronic continuity of the normative functional roles of Manifest-image concepts, both of which account for their persistence in the face of scientific advances.

The present collection concludes with a transcript of the author-meets-critics session that capped off the conference. The session, chaired by Michael Hicks, began with supplementary remarks by Brandom, which provided further clarification of his critical interpretive project. O'Shea and deVries then offered extensive commentaries on Brandom's arguments, both for his positive claim about the pragmatically metalinguistic character of framework-structuring categorial concepts, and for the negative claim that Sellars' adaptation of Kant's phenomenal-noumenal distinction is untenable. Brandom's brief replies to his critics were followed by a Q&A period, during which audience members participated in a lively discussion. The way in which the discussants circled around particular questions, coming back to earlier comments and extending or elucidating them, strikes us as an exemplar of the philosophical process at its best. In editing, we viewed the transcript as an historical document that would provide not only a snapshot of Sellars scholarship in 2015, but also some insight into what the tenor of the discussion was like—an engagement with complex topics, punctuated with humor and off-the-cuff remarks. For the reader's benefit, we retained the jokes, asides, and other quirks of the conversation.

As the above survey indicates, the contributions to the present volume open a number of new avenues of research, and provide ample opportunity for young scholars to adapt Sellars' philosophy to new domains. Our hope is that the next generation of Sellars scholars will draw on this collection in effecting a deeper understanding of agency and practical reasoning, semantics and pragmatics, speech acts and mental events, subpersonal representations, sensorimotor skills, expressivism, deflationism, the ontology of colors and hammers, the rational constraints on perception, the role of framework-structuring concepts, the modal constraints on cross-sortal identities, and ultimately, the relations between the Manifest and Scientific images. These lines of investigation will be at the vanguard of Sellars scholarship in the years to come.

David Pereplyotchik
Deborah R. Barnbaum
Department of Philosophy
Kent State University

References and Further Reading

Brandom, R. B. (2002). *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality.* Cambridge, MA: Harvard University Press.

Brandom, R. B. (2008). *Between Saying and Doing*. New York, NY: Oxford University Press.

Brandom, R. B. (2015). *From Empiricism to Expressivism: Brandom Reads Sellars*. Cambridge, MA: Harvard University Press.

Chisholm, R. and Sellars, W. S. (1957). "Intentionality and the Mental," *Minnesota Studies in the Philosophy of Science*, 2, 507–39.

Coates, P. (2007). *The Metaphysics of Perception: Wilfrid Sellars, Perceptual Consciousness and Critical Realism*. New York, NY: Routledge.

deVries, W. A. (2005). *Wilfrid Sellars*. Philosophy Now Series, Chesham, Bucks: Acumen Publishing and Montreal and Kingston: McGill-Queen's University Press.

deVries, W. A., ed. (2009). *Empiricism, Perception, Knowledge, Normativity, and Reason: Essays on Wilfrid Sellars*. Oxford: Oxford University Press.

Olen, P. (2016). *Wilfrid Sellars and the Foundations of Normativity*. London: Palgrave Macmillan.

O'Shea, J. R. (2007). *Wilfrid Sellars: Naturalism with a Normative Turn*. Cambridge, UK and Malden, MA: Polity Press.

O'Shea, J. R. (2014). *Kant's Critique of Pure Reason: An Introduction and Interpretation*. New York: Routledge.

O'Shea, J. R., ed. (2016). *Wilfrid Sellars and His Legacy*. Oxford: Oxford University Press.

Price, H. (2011). *Naturalism without Mirrors*. New York: Oxford University Press.

Rosenberg, J. (2007). *Fusing the Images*. Oxford, UK: Oxford University Press.

Rosenthal, D. M. (2005). *Consciousness and Mind*. Oxford: Oxford University Press.

Sellars, W. S. (1956). "Empiricism and the Philosophy of Mind," *Minnesota Studies in the Philosophy of Science*, vol. I, H. Feigl and M. Scriven (eds.), Minneapolis: University of Minnesota Press: pp. 253–329. Reprinted in *Science, Perception, and Reality*.

Sellars, W. S. (1963). *Science, Perception, and Reality*. London: Routledge & Kegan Paul Ltd and New York: The Humanities Press, 1963; reissued in 1991 by Ridgeview Publishing Co., Atascadero, CA.

Sellars, W. S. (1967). *Science and Metaphysics: Variations on Kantian Themes*. Atascadero, CA: Ridgeview Publishing Company.

Sellars, W. S. (1972) "' . . . this I or he or it (the thing) which thinks . . . '," *Proceedings of the American Philosophical Association*, 44, 5–31.

Part I

Ethics, Moral Reasoning, and Free Will

1 Thought, Freedom, and Embodiment in Kant and Sellars

James R. O'Shea

In the preface to *Science and Metaphysics,* Sellars remarked on the "astonishing extent to which in ethics as well as in epistemology and metaphysics the fundamental themes of Kant's philosophy contain the truth of the variations we now hear on every side" (SM x).[1] Also astonishing, in many ways, was Sellars' own 1970 Presidential Address to the *American Philosophical Association* (APA), which borrowed its title from the phrase in Kant's Paralogisms, ". . . this I or he or it (the thing) which thinks . . ." (A346/B404).[2] In its compact twenty-five pages Sellars managed to sketch novel yet plausible reconstructions of central aspects of Kant's views on self-knowledge, persons, freedom, and morality, along the way suggesting how all of those Kantian views could plausibly be rendered consistent with a naturalistic ontology. In this chapter I want to focus on Sellars' APA address as an occasion for reflection on how both Kant and Sellars—or so I will contend—offer insights into how we ought best to conceive the nature of and the relationships between our thinking selves, our practical agency, and our entirely natural, material embodiment.

I

On Sellars' view, and I think for Kant too, what most fundamentally unites the theoretical domain of our cognition of objects with the practical domain of freedom and morality is our capacity for *thought*; or more particularly, for thought as governed by various norms of reasoning. Both sense perception and volition, for example, are fundamentally species of conceptual *thinking*, for Sellars. As he put it in his 1967 Lindley Lecture on ethical theory:

> (§1)The focal point of practical reasoning is action, as the focal point of empirical reasoning is observation. Perceptual takings or 'judgments' are the thoughts which typically arise from the impact of the world on our mind through our sensory capacities. Volitions are the thoughts which typically impinge on the world through our motor capacities.
>
> (§2) Intentions can be thought of, somewhat metaphorically, as practical commitments. Volitions can correspondingly be thought of as

> practical commitments to do something *here* and *now*, and hence as a special case of commitments to do something at sometime or other.
>
> (FCET §§1–2)

Sellars' primary *model* for thought—that is, for how to understand what thought *is*—is in terms of our public verbal behavior as reflected in various norm-governed causal patterns of inference, and in various norm-governed causal patterns of perceptual response and intentional action in relation to one's environment. Sellars furthermore famously argued in 'Empiricism and the Philosophy of Mind' (EPM 1956), in his 'myth of genius Jones' (cf. EPM X–XV), that the further naturalistic posit of a kind of representationalist "Mentalese" or language of thought in the mind/brain was not only permissible but required on overall explanatory grounds. But for present purposes, we can stick to the public, pragmatic, and Wittgensteinian features of Sellars' conceptions of meaning and of thought in terms of norm-governed usage and of functional roles respectively, which for most philosophical purposes also fits the spirit of Kant's view of concepts as involving in the prescription of rules (cf. O'Shea 2016a).

In his APA address on Kant, accordingly, Sellars focuses on the character of our *thinking* across both the theoretical and practical domains. The lead question of the article is the same as Kant's in the Paralogisms: namely, what is the nature of our thinking being, or as Sellars puts it, of 'an I', or of 'this I', "meaning roughly whatever can be referred to by an appropriate tokening of 'I'" (I §3)? He then proceeds to defend both Kant's epistemic account of the *a priori* unity of an I, as a necessary condition of the possibility of experience, and Kant's diagnosis of the fallacies involved in traditional metaphysical accounts of the nature of such a thinking being, as the subject matter of a putative 'Rational Psychology' (I §5). Here I want to focus on the ways in which the Kant-Sellars view of the 'I think' relates to the wider systematic issues pertaining to nature, freedom, and morality that Sellars discusses in the rest of the APA article.[3] Along the way I will argue for modifications to the views of both Kant and Sellars such that in the end a distinctively *Kantian naturalist* outlook on these issues is supposed to emerge as plausible, at least in broad strokes.

The most abstract but also the most fundamental unity of the thinking self, for Kant and Sellars, is what might be called a *thought-unity*: a conceptually represented unity, but one that consists in the unity of a form of representation rather than the representation of the self as a unified *thing* or object or substance, whether material or immaterial. Kant's arguments for this are familiar but controversial. They start out, for example, with the distinction between manifolds or successions of thoughts or experiences taking place *in* a consciousness, in contrast to the thought or experience *of* such manifolds or successions *as* a manifold or a succession, in a judgment or conceptual rule. The latter sorts of object-constituting thought-unities or conceptual rules are argued to be

necessary for the possibility of any potentially self-aware experience—for example, for the possibility of any awareness or representation of oneself *as having* experiences of a world of objects or phenomena that exist independently of those experiences. On this Kant-Sellars view, a thinking self is, at least with respect to this first most abstract necessary condition, itself an achievement of representation or thought. The key point is this: on this view, it *just is* in thinking in accordance with the rule-governed conceptual distinctions and experience-informed inferential patterns that implicitly distinguish a generally lawful world of objects from one's perspectival experiences of them, that we thereby implicitly represent *ourselves* as experiencers of that world in the first place. On Kant's view, to uncover this feature of our consciousness through transcendental abstraction or reflection is not to reify the self in any further sense than as just stated. This is why, for example, if one goes *searching* for the self phenomenologically or experientially, one finds it to be systematically elusive (cf. O'Shea 2015).

What Sellars rightly goes on to emphasize in this account, however, is what it does *not* entail, both what it rules out and what it leaves open. By contrast, what I will call the general *disembodiment* style of objections to Kant's views on these and other matters, both in his theoretical and his practical philosophy, is the familiar form of criticism that in thus conceiving the self (or as we shall also see, freedom and morality) in such attenuated, abstract, or formal terms—and worse, of course, as also characterized by apparent assertions of ultimately "supersensible" or non-spatio-temporal existence—Kant renders impossible any ultimately satisfactory account of our naturally embodied cognition and practical agency in the empirical world. However, I think that in cases where there is something right about the disembodiment style of objections to Kant's view, at least as far as Kant himself is concerned, more often than not the objections locate the problems in the wrong place. And unlike those more familiar objections, in the APA paper Sellars sketched original interpretations of Kant's theoretical and practical philosophy that, while departing from Kant in certain key respects, provide what I contend is an unusually penetrating interpretation of Kant's own views.

The present case of the nature of the thinking self is a particularly difficult one to sort out in this respect. Sellars begins by analyzing Kant's view of the I in the Paralogisms as supporting three main points:

(1) the I is a being of unknown species which thinks;
(2) the I doesn't simply 'have thoughts': *it thinks*—but in knowing *that* it thinks, and *what* it thinks, we are not knowing what sort of being it is.

Yet,

(3) the I must have a nature—what it is we cannot know, though we *can* know that it is not material substance.

(I §22)

Sellars explains Kant's claims roughly as follows. What we know *a priori* is the I as a form of thought-unity, as I have put it. We cannot, by contrast, know *a priori* the nature of the I as a thing or object (i.e., as the 'ultimate subject' of our thoughts in *that* sense). Kant's further claim, that we *can*, however, know that the nature of the I, is *not* material, pertains to matter conceived as the movable in space, as *res extensa* (or more accurately, as fields of force) in space and time.

The ground of claim (3) is not, Sellars says, because on Kant's view we have any "positive, let alone adequate, idea of mind as a *sort of being*" (I §18). Rather, it is in part because of the arguments mentioned above from the Deduction. The materialists and reductive empiricists are in danger of missing the crucial transcendental point that the *irreducible representational role* of the 'the simple I' of apperception cannot be accounted for in terms of concepts pertaining to the spatial pluralities with which matter presents us. And that is correct, I think, insofar as we are talking, at least at this level of functional abstraction, only about the necessary unity of a certain form of representation or thought-unity. Furthermore, Sellars suggests, (3) finds additional support for Kant from his view that our empirical knowledge of the self finds no space, so to speak, in the primary qualities of matter for our mental states as we are aware of them in introspection or 'inner sense'.[4] In the background here is Kant's general Galilean view that the secondary sensible qualities, of color as experienced, for example, must have their empirical home somehow in the experiencer in contrast to the fields of force and materials that constitute *space*, for Kant. Kant himself is in the end, in this domain, a kind of agnostic inner/outer *empirical dualist*: as Sellars puts it, instead "of opting for a Strawsonian account of the empirical self," that is, a neo-Aristotelian account of selves as embodied persons with both mental capacities and physical characteristics, "Kant opts for a dualistic model" (I §41; cf. §18).

Again, he does not do so because we have any positive idea of what a mind is *per se*; rather, we know only the represented unity in our thinking that is achieved by our own thoughts, as noted above. If we move from consideration of the sensory aspects of consciousness to our conceptual thinkings *per se*, these for Kant are primarily functional-representational activities or rules of unity (cf. I §§20–1), consisting for example in the representation of law-governed modal constraints in experience. For the Critical as opposed to pre-Critical Kant, our thinkings do not, contrary to Descartes and other Rational Psychologists, present themselves to us as having or entailing a *nature per se* (as attributes of an immaterial "mental substance"; cf. I §§18–19).

We might add that grounds for Kant's denial of materialism in claim (3) can be found in his reasoned opposition throughout the Transcendental Dialectic to any global ontological claims about the nature of the cosmos as a complete whole *per se*. For Kant the latter indirect support for his transcendental idealism entails that all *denials* of immaterialism and of theism

(as represented in the ideas of soul and God respectively) are as dogmatic as their speculative assertion, though the ideas themselves are warranted for us merely as regulative 'as if' ideas that guide our practical hopes and our unending theoretical enquiries (cf. O'Shea 2014, Chs. 2 and 6). With reason's regulative idea of the soul, according to Kant, we can and should combat any globally assertive materialistic naturalism by thinking of our own thinking nature *as if* it were a non-spatial unified being or substance. But for Kant no such conclusion can in fact be non-fallaciously inferred from anything we actually know about our thinking selves.

So our knowledge of the I, for Kant, is limited in principle to the various *a priori* and empirical resources described above. The texts of Kant's Paralogisms do broadly support Sellars' main interpretive moves in this domain, or so I would claim.

Sellars begins to bring out the primary *revision* that he thinks needs to be made to Kant's view in this case—and this, I think, is one element of truth in the familiar disembodiment objections—by pointing out certain ontological possibilities that Kant in the Paralogisms recognized as at least logically thinkable. Sellars quotes from the following passage:

> The identity of the consciousness of Myself in different times is therefore only a formal condition of my thoughts and their connection, but it does not prove at all the numerical identity of my subject, in which—despite the logical identity of the I—a change can go on that does not allow it to keep its identity; and this even though all the while the identical-sounding "I" is assigned to it, which in every other state, even in the replacement of the subject, still keeps in view the thought of the previous subject, and thus could also pass it along to the following one.
>
> (A363)

And he comments on this as follows:

> [Kant] is suggesting that the "logical identity" of the I through Time, which is an analytic implication of the knowledge of oneself as thinking different thoughts at different times, is compatible with the idea that these thoughts are successive states of different ultimate subjects. Compare the materialist who argues that the thoughts which make up the history of an I are states of systems of material particles which are constantly losing old and gaining new constituents.
>
> (I §27)

I will comment on Sellars' final remark about materialism further below. For the present, the point is that it is at least barely thinkable that *the mind*—that is, the thinking self as thought-unity—might not turn out to be, in itself, an *ultimate logical subject* at all, but rather a series or plurality of ultimate subjects. In a different passage (A359), Kant also suggests the

thinkable possibility that the thought-unity of the thinking self might turn out to be, 'in itself' (i.e., as 'noumenon'), a unified substantial *person* as ground of both one's mental capacities and one's spatial-material attributes. Kant's Paralogisms thus contain thought experiments according to which such an ultimate logical subject could at least thinkably be, in itself, either a substantial unity or an ultimate plurality of substances (with memories passed from one 'substance' to another, for example).[5]

But Sellars' main point here is this: Kant held that the thinking self or the I is *knowable* primarily only as a set of *capacities* of various kinds, and for Kant we must remain agnostic about the ultimate grounding or realization of these capacities. What Sellars himself is proposing at this point is easily misunderstood, however, especially given his comparison with "the materialist" quoted above. Sellars' argument is that Kant's own view of the thinking self, in the specific respects just noted, has shown why *we* can make a different move from Kant at this point (i.e., why it is intelligible *for us* to reject the agnosticism inherent in Kant's own *empirical dualism* of the inner mental versus the outer material, while nonetheless maintaining Kant's conception of the thought-unity of the thinking self). We can then intelligibly defend a view, which Sellars is clear that Kant himself did not hold, of our knowable *empirical* selves as fully materially embodied *persons* exercising various normatively characterized mental abilities and possessing various physical attributes. For what Kant has argued to be at least thinkable is that a logical subject that is necessarily not *represented as* an aspect of something more basic could nonetheless, in itself, turn out to be identical with something that has both thinking capacities and material attributes. And so we, if not Kant, can argue that in fact our irreducibly normative thinking capacities are fully realized in, and so (as Sellars puts this suggested revision of Kant) can in fact be "*identical with*[,] the being which, as having material attributes, is the body" (I §30).

I think both the interpretive and the reconstructive aspects of these contentions of Sellars' are basically on target here too, although Sellars' often unclear dialectical qualifications in the APA paper seem to have led some Kant interpreters to think that Sellars was basically or indirectly attributing our *materialist* functionalism *to Kant*, which as we have seen is not what Sellars was doing.[6] The particular speculative suggestion is not that *Kant himself* was either a functionalist or a non-reductive materialist but that *we* might adopt and put Kant's own critical-transcendental account of the thinking self to use within what we might call a *Kantian naturalist* conception of our irreducibly normative yet fully materially embodied thinking nature.

From this perspective we can agree with Kant's famous argument in the Refutation of Idealism, while rejecting Kant's own inner/outer empirical dualism. That is, we can agree that the cognition of the temporal ordering of one's own mental states depends upon a prior background involving the direct cognition of persisting material objects in space in general. But what

Sellars notes now, however, is that given Kant's view of the nature of the world of appearances in space and time, presumably the mental states of our reconstructed, fully embodied (Strawsonian-Aristotelian) persons, just as much as is the case in relation to Kant's own empirical dualism of the objects of inner and outer sense, "would be 'appearances' belonging to a deterministic natural order" (I §40). For as Sellars notes, all of the "states of the empirical self," for Kant, both inner and outer, "belong to a deterministic system of events the core of which consists of material events occurring to interacting material substances" (I §44). The problem of determinism thus provides Sellars with a quiet segue into the second half of the APA paper, on topics concerning freedom and morality.

What eventually becomes clear by the end of the paper is that Sellars takes Kant's reference not just to "this I or he" but in particular to "*it, the thing*" that thinks, to be a hint that all of the preceding views about the irreducible yet (we might hold) fully materially embodied thinking self has not yet by itself given us the human *person*, properly speaking (cf. I §62–3). Rather, by themselves the sorts of theoretical considerations (in Kant's sense) considered so far might conceivably only give us what Sellars calls "an *automaton spirituale* or *cogitans*, a thinking mechanism" (I §63).[7] As he puts it:

> What is haunting Kant, in this cryptic passage [i.e., "this I, or he or it, (the thing) which thinks"], is the concept of an *automaton spirituale*, a mind which conceptualizes, but only in response to challenges from without, and in ways which, however varied, realize set dispositions.
>
> (I §65)

With regard to "it, (the thing) which thinks" Sellars refers us to Kant's statement in his later *Metaphysics of Morals* that a "*thing* is that to which nothing can be imputed" (MM 6:223).[8] And if we follow up that reference further, we find that Kant precedes that statement about thinghood with this one:

> A *person* is a subject whose actions can be *imputed* to him. *Moral* personality is therefore nothing other than the freedom of a rational being under moral laws (whereas psychological personality is merely the ability to be conscious of one's identity in different conditions of one's existence).
>
> (Kant MM 6:223)

In roughly the second half of the APA paper, then, what we find Sellars wrestling with is the question of how practical reason and intentional action, and hence freedom and morality, are related to our embodied personality as thinkers, or as 'thought unities', discussed above.

Here in the practical domain, as we saw earlier in the theoretical domain, there will be further opportunities for understandably misinterpreting Sellars' carefully qualified reading of Kant. We have seen above that Sellars was not attributing to Kant an implausibly prescient materialist-functionalist philosophy of mind, but rather an empirical dualism—one that was, however, grounded in revolutionary formal insights about concepts as functions and about the unity of thought, which Sellars argues *we* can put to Strawsonian and functionalist uses. Similarly here, Sellars is not suggesting, for example, that Kant himself *did* conceive the unity of thought in the theoretical domain in accordance with an *automaton* conception of theoretical thinking, one that is characterized by what Sellars calls a merely "*relative* spontaneity," and which would then be sharply contrasted with the genuine spontaneity or autonomy of transcendental freedom. There is truth in the contrast between the two, but not when reified in that way. In my view, Sellars sees that one important key to Kant's philosophy in both domains is that, for Kant, abstraction in these respects does not entail reification. That is, Kant's method involves abstracting from our embodied experiences and actions various formal principles of thought and freedom that are thereby revealed as having necessarily been operative within those embodied realities themselves, without this reflective distinction by itself entailing any real disembodiment or non-spatio-temporal thinghood *per se*, whether known or unknown. What Sellars saw more clearly than most Kant commentators, I believe, is that not only embodiment but also a primacy accorded to *practice* and *purpose* is implicit throughout the rarefied non-empirical atmosphere of each of Kant's three highly abstract *Critiques*, including the Transcendental Analytic of the first *Critique*.

On the theme of determinism, Sellars begins by commenting as follows on Kant's view of the *self-affection* that is involved in the mind's own introspective searching of its own contents (analogous to our being affected by outer objects):

> §66. [Kant] grants that 'inner perception' may be prepared for by an activity of searching, a direction of attention in which the mind affects itself, just as perceptual response may occur in a context in which we are looking for something, seeking relevant observations. But why the direction of attention? Relevance to what? Here considerations of purpose enter in and the first *Critique* simply *abstracts* from the purposive aspects of the conceptualization involved in experiential knowledge.
>
> §67. Now it is clear that although the structure of the first *Critique* highlights what I have called the relative spontaneity of the conceptualizing mind, it clearly presupposes a larger context in which the mind is thinking to some purpose. Thus reference to reason in its practical aspect is implicit throughout the *Critique*, but only in the Dialectic, after the constructive argument is over, does it become explicit.
>
> (I §§66–7)

I think the outlook sketched in these passages is crucial for correctly interpreting Kant's critical philosophy across the board, and in particular for questions concerning how to understand Kant's *applications* of his transcendental principles in his writings beyond the three *Critiques*.[9] Those applications involve increasingly empirically contentful and correspondingly *a priori* indeterminate and merely regulative domains—applications that often, not surprisingly, have thus seemed to many interpreters to involve inconsistent shifts in Kant's doctrines across these different contexts. But Kant's regulative maxims of reason and reflective judgment, for instance, are not *solely* heuristic 'as if' principles of organization and search. More fundamentally, Kant takes these more empirical and (as he puts it) "indeterminate" regulative principles and generic material applications to be necessary for the very possibility of the sorts of determinative categorical experience that he had already analyzed, with respect to its most abstract structures, in the Transcendental Analytic of the first *Critique*. Kant says this explicitly in various places, but it is understandably difficult to keep what is all along supposed to be kept implicitly in mind as one moves along from Kant's Analogies of Experience, through his regulative maxims of reason, to the regulative principles of purposiveness in the Third *Critique* and beyond to all their various intended applications in nature and in action.

II

So, what *does* happen in the second half of Sellars' APA paper, by way of making explicit the wider practical context that was left implicit in the above doctrines concerning our embodied conceptual thinking? We left off at the point where Sellars had noted that the "states of the empirical self" for Kant, both inner and outer, "belong to a deterministic system of events" (I §44). Here Sellars begins by introducing a further modification of Kant's views concerning the empirical self. He suggests that Kant, like many other philosophers, implicitly assumes that if nature is indeed a deterministic system, it must follow from this that we ourselves are *passively caused to be in* whatever mental or physical states we are in. "The picture," Sellars writes, "is that all natural objects are passive with respect to their states—so that if they cause other things to change, they do so because they have, in their turn, been caused by other things to be in the state by virtue of which they are causes" (I §48).

By contrast, Sellars contends (and argues elsewhere) that "*The past is not something with respect to which we are passive*" (I §49), and that this is so even on the assumption that we along with everything else are thoroughly ensconced within nature considered as a deterministic, physical system. This idea reflects Sellars' own unique version, at least in this respect, of the familiar compatibilist contention that only in certain kinds of circumstances are we (as we would ordinarily say) *forced* or compelled by "foreign causes," as Kant puts it (cf. I §50), to think or to act or to will as we do. In our thinking, for example—and hence in our intentions, volitions, and actions, too—we

are not entirely the playthings of nature but are also actively thinking, self-monitoring systems, as it were. "Pure apperception," writes Sellars, "gives us a non-passive awareness of the mind as active. Indeed [he continues], Kant insists again and again that the mind is aware of the 'unity' and 'spontaneity' of its acts of synthesis" (I §56).

But Sellars does argue plausibly that, for Kant, the spontaneity of which we are thus aware, considered so far solely from the theoretical perspective, and in relation to the determined appearances of inner and outer sense, might, for all that, still be the merely responsive "relative spontaneity" of a "thinking mechanism" (I §63)—a "*theoretical automaton spirituale*" (I §68), as Sellars puts it. Drawing the familiar computer analogy, he suggests that in this sense the mind's spontaneously initiated logical 'searches' in response to data and given its own "computational dispositions" (I §58) would ultimately still be *a* form of passivity, "though not *sheer* passivity" (I §59). The relative spontaneity of pure apperception, so considered, would in this case remain, as Sellars puts it using Kant's terms, "another example of a cause the causality of which is [itself] caused" (I §59).

It is only in the final quarter of the APA paper that Sellars comes to the properly practical domain from which the earlier analysis of our self-conscious theoretical cognition has abstracted. This concluding analysis takes place in two Kantian steps: not surprisingly, first in relation to what would basically be a neo-Humean conception of *heteronymous* agency, in Kant's sense; and then finally in relation to *autonomous* agency, or "acting for the sake of principle" (I §85): "freedom in the deeper sense . . . [Kant] is seeking to explicate."

First, then, Sellars explains that the relatively spontaneous "freedom of choice" or *Willkür* of (in this case) a heteronymous agent would essentially involve the agent's combining various intentions and desires with factual information in a procedure that generates "alternative courses of action, one or other of which, *ratified* by the appetitive faculty, would become the decided course of action" (I §68). Within this picture is a higher-order practical premise of "self-love," to use the traditional term, reflecting the natural human interest in promoting one's own happiness. Practical reason, *exclusively* so conceived, would not have a principle that is peculiar or *intrinsic* to itself, as Sellars puts it; rather, our practical reason would serve, in the above way, the naturally implanted end of pursuing one's own happiness, for example, in relation to one's other "particular desires and aversions" (I §75).[10]

By contrast, second, Sellars cuts quickly to the chase scene by bluntly formulating his own version of the practical premise that is intrinsic to practical reason, what Kant calls 'pure' practical reason, in the following way (Sellars develops this elsewhere, in particular in SM chapter 7; see Koons, this volume, for further discussion):

> [The *practical premise* that is intrinsic to practical reasoning is:] "Let any of us persons do that in each circumstance which promotes our common good."
>
> (I §79)

Sellars indicates that he will "not attempt to justify the ascription of exactly this premise to Kant," though he does remark that Kant's own fundamental moral law is *generic* in character rather than "purely formal" as interpreters of Kant, he thinks, have mistakenly supposed. It does seem to me that scholarship on Kant's practical philosophy over the last several decades has indeed increasingly stressed that Kant's moral law, for all its formal universality, is reflectively abstracted from the ways in which it functions both within the natural pursuit of human happiness and in the cultivation of a virtuous character, and hence in relation to our naturally embodied feelings and inclinations. The case in this respect closely parallels the points I stressed earlier in relation to our materially embodied capacity for 'pure apperception' as thinkers, as well as concerning the intended empirical realization of Kant's transcendental principles as one moves from the first two *Critiques* to the third *Critique* on natural purposiveness and beyond to Kant's later writings. In the Presidential Address, in my view, Sellars' remarkable if unheralded achievement was to have articulated the most fundamental abstract principles of the first two *Critiques* in a way that, against the tide of Kant interpretation at the time, correctly situated Kant's accounts of cognition and freedom in relation to their intended naturally embodied realizations.

This is not the occasion to attempt an explication of Sellars' moral theory in detail, but for present purposes a few words on the core truth that he finds in his reformulation of Kant's moral law will be helpful. The moral point of view consists most fundamentally in our capacity to have practically efficacious thoughts, the content of which, as generically expressed in the above principle or "practical premise," is a shared intention or practical commitment that, crudely put, any of us persons act in ways that promote our common good. This of course immediately generates a host of questions concerning the specification of the relevant 'we' or community of 'persons', and concerning the idea that, as Sellars claims, this practical premise "constitutes a purpose which can be said to be implied by the very concept of a community of persons" (§79). And Sellars also recognizes that the compact generic reference within the Kantian practical premise to what, in each circumstance, "promotes our common good," glosses over empirical realities and uncertain means-end reasonings of enormous complexity. Without ignoring the existence of those and many other difficulties, however, the following core conception of our practical agency emerges in Sellars' closing discussions.

Whatever might thus be taken to be the means to promote our common good, Sellars here simply dubs "condition α," so that the generic practical premise now becomes: "Let any of us persons do that which satisfies condition α," from which premise, along with relevant information, practical reasoning would derive the particular volitional conclusion: "Let me now do *A*" (§80). Importantly, however, that I actually *do A*, Sellars points out, assumes not simply that I understand that implication (i.e., the implication, roughly: "Let any of us do α, so let me now do *A*"), but also that I actually

affirm the antecedent of this implication (i.e., the generic practical premise itself). We can see what Sellars has in mind here when he brings in the typically conflicting practical premise of self-love (or of sympathy, or of some other fundamental first-order interest). For then "in cool hours," as Sellars puts it echoing Joseph Butler, we will confront within ourselves the following two opposed practical thoughts:

> (1) "Let me now do *A*, *because* let any of us do actions satisfying α, *although* this implies not promoting my happiness by doing *B*." [This would be acting autonomously "from the moral point of view", i.e. "choosing to do something for the reason that it is implied by the moral law," that is, "as being what I ought to do" (I §§83–5).]

As opposed to:

> (2) "Let me now do *B*, *because* let me now promote my happiness, *although* this implies my not doing *A*, which is subsumable under the principle 'let any of us do actions satisfying α." [This is to "choose" (*Willkür*) from "the 'personal' point of view" (I §82), whether from self-love, as here, or from sympathy or from any other feeling or interest.]

The moral point of view is thus a form of *thinking*, for Sellars' Kant: an internally motivating thought or efficacious intention (once "affirmed") that is *generic* in its content, as being a commitment that any of us persons perform certain kinds of action in certain kinds of circumstances. As Sellars puts it: "That practical reason is autonomous means that a choice is possible in which practical reason itself affirms the antecedent" (i.e., the generic practical premise, rather than one's choice issuing solely from the relative practical spontaneity that is ultimately, however complex, a so-called ratification by one's inclinations or interests [I §85]). It is our capacity to act autonomously "for the sake of principle," Sellars concludes, that *distinguishes us from what would otherwise, without that capacity, be the exclusively heteronymous choices of an "it (the thing) which thinks."*

Note that Sellars has here quietly exploited Kant's distinction between *Willkür*, or freedom of choice, as distinguished from an autonomous *Wille* as the generic or abstract motivating intention to act from the moral point of view (I §§81, 85). Supposing we lacked an autonomous *Wille*, or an intrinsically practical reason in Sellars' sense, he points out that we would still exercise the relatively spontaneous but ultimately heteronymous power of appetite-ratified choice as outlined above. But crucially, Sellars' particular account explains how it is that, even as the autonomous rational beings that we are, we must still always exercise choice or *Willkür* as to the matter of *which antecedent premise-intention is (implicitly or explicitly) affirmed in the circumstances* (i.e., that of morality or, for example self-love). For Sellars is clear that it is *not* intrinsic to autonomous practical reason, but rather

a matter of choice (*Willkür*), that one in point of fact affirms one rather than the other—that is, that the one motivating thought rather than the other motivating thought in point of fact turns out to be efficacious in motivating one's action. (Though Sellars does not mention it here, this account is presumably also designed to help with the classic interpretive problem of addressing how Kant's moral philosophy is able to account for our freedom to choose to act *immorally*—that is, to freely choose evil.)

In closing his address, Sellars adds these final remarks on the wider picture that involves Kant's *transcendental idealism* (I briefly come back to this in part III, sections [3] and [4] below):

> §86. Kant ends on an agnostic note. We are conscious, in pure apperception, of ourselves as autonomous rational beings, beings which *can* act out of respect for principle. But is not, perhaps, this consciousness an illusion? He claims to know, on philosophical grounds, that as objects of empirical knowledge we are *not* autonomous beings. We cannot, alas, show, on philosophical grounds, that *as noumena* we *are* autonomous. He therefore takes refuge in the claim that, equally, we cannot *know*, on philosophical grounds, that as noumena we are *not* autonomous.
>
> (I §86)

I will close with some final reflections on this Kant-Sellars conception of thought and of agency as it has emerged in striking form in Sellars' APA address.

III

[1] First, note Sellars' remark, just quoted, that "we are conscious, in pure apperception, of ourselves as autonomous rational beings." This remark restores the implicit practical context of autonomy from which the theoretical examination of the merely "relative spontaneity" of pure apperception had abstracted for the analytical purposes then at hand. This is not to say that our theoretical judgments and 'syntheses', according to either Sellars or Kant, are *acts* in the sense of deliberate, intentional *actions*. But it does suggest that both conceptual thinking and practical thinking for Kant and Sellars involve the capacity to reflect on principles, and also that without the implicit wider practical and purposive context we would not be the self-consciously apperceptive thinkers that we are.

[2] Second, it seems to me that there is a complex, at least ostensible circularity implicit in Sellars' account of the role of normative principles in our practical agency. The situation, I think, closely parallels the ostensible circularity that Sellars wrestled with throughout his career in the theoretical or epistemic domain: namely, that between our general epistemic principles on the one hand, and the particular non-inferential judgments that such

principles are supposed to warrant on the other. Elsewhere I discuss what I call Sellars' *Kantian naturalist* solution to that epistemic circularity problem (most recently in O'Shea 2016b, section 4), and the parallel issue in the present case can be stated this way. The conception of normative 'oughts' that Sellars has exploited in his discussion of agency in the APA paper rests elsewhere on his more general account of 'we intentions' (e.g., cf. SM chapter 7 and Koons, this volume).[11] But the latter account of community intentions rests on Sellars' conception of individual intentions as practically efficacious motivating thoughts that have both their sense and their efficacy only within an *already* normative 'space of reasons' (i.e., within a framework of implicit '*ought-to-be*' rules governing our linguistic and other behaviors). Put more curtly, then, Sellars explains normative principles or 'oughts' ultimately in terms of certain kinds of intentions; but what it is to be an intention, on his view, is *constituted* by its role within an already normative space of reasons and linguistic *ought-to-be*s. The ostensible circle has '*oughts* explained in terms of intentions' and yet 'intentions explained in terms of a space of *oughts*'.

I think Sellars' way of seeing this ostensible circle as benign or merely apparent will have something like the same structure as the Kantian naturalist account of the warrant for epistemic principles that he gives, for example, in "More on Givenness and Explanatory Coherence" (MGEC). Along one dimension, that is, normative principles are 'always already there' in our cultural inheritance in its continuing conceptual and practical development over time. From this perspective it is norms both 'all the way down' and 'all the way back', so to speak. Within this outlook a kind of diachronic reflective equilibrium is continually sought between our particular sets of judgments, inferences, and actions, and the more generic but revisable normative principles that are espoused and reflected in those practices. This dimension also includes for Sellars, both in the APA article and in MGEC, various stronger Kantian-style claims: for example in the practical domain, as we have seen, claims about personhood as entailing a certain autonomous capacity to institute norms and to act on principle as such; and claims about moral commitments as in some way conceptually connected to the idea of a *community* as such.

But in the epistemic domain, at any rate, Sellars indicates that there must also be another, circle-breaking *naturalistic* dimension of explanation as well, and in particular a biological evolutionary explanation of the ultimate origin of our capacity for conceptual-linguistic *thinking* itself. As Sellars wrote in MGEC:

> Presumably the question 'How did we get into the [epistemic] framework' has a causal answer, a special application of evolutionary theory to the emergence of beings capable of conceptually representing the world of which they have come to be a part.
>
> (MGEC §79)

Nature in itself, for Sellars, in one primary sense, knows no 'oughts'. But natural selection over time, according to Sellars, generates patterns of bodily-environmental, instinctual-behavioral, and proto-cognitive adaptation such that what it is to be any such item so functioning cannot be understood apart from its place within such a wider normative network or 'selection space' of biological ought-to-bes. The Kant-Sellars account in the APA paper of both our apperceptive intelligence and our autonomous agency must be understood as all along embedded within a wider framework of our *naturally* adaptive or purposive animal inheritance. In a structurally similar but of course altogether differently grounded way, the highly abstract 'analytic' sections of Kant's first and second *Critiques* arguably cannot be understood except in terms of their always intended embodied embedding within the regulatively systematic conception of *natural purposiveness* that Kant outlines in the appendix to the Dialectic in the first *Critique* and then subsequently expands upon in the third *Critique*.

[3] Third, we must in the end come to terms with the elephant in the room: Kant's transcendental idealism, and Sellars' proposal to replace Kant's *noumena* with the postulation of unobservables that is characteristic of ongoing scientific theorizing. I have been focusing on Sellars' naturalism conceived as integrated fully *within* the above Kant-Sellars story in the APA address about our embodied practical agency in the knowable material world, as supplemented with some amplification of Sellars' conception of our evolved animal purposiveness. I have not discussed Sellars' *own* quasi-Kantian transcendental idealism, with its *scientifically* reconceived noumena put forward as a replacement for Kant's unknowable 'things in themselves', and with the object-ontology of the manifest image, with respect to that 'noumenal' object-ontology, supposed to be revealed to be a framework of *mere* 'appearances' that is *strictly speaking* false (for the details, see O'Shea 2016b; for a critical assessment of Sellars' strategy here, see Brandom, this volume).[12] Sellars was well aware that Kant himself did not defend the particular 'analogical' and 'relocational' story about our sensory consciousness of colored objects, and so on, that forms the centerpiece of Sellars' own confessedly non-Kantian argument for transcendental idealism in chapter 2 of *Science and Metaphysics*. Kant's own transcendental idealism was concerned primarily with recommending the wholesale *critical replacement* of the traditional object-oriented ends of speculative theoretical metaphysics with the idealized ends of our own practical rationality and moral freedom.

On Kant's view we *must* in the end think of nature analogically in terms of an all-powerful intelligent designer's purposive practical handiwork, thus experiencing nature 'as if' it were designed for us to know it through our contingent empirical researches, and 'as if' it were built for the gradual achievement of our highest good using the crooked timber of our practical agency. Kant himself thus formulated our inevitable analogical thoughts about unconditioned 'things in themselves' in terms of the theoretical and practical regulative *ideas* of God and of our own souls, conceived

by analogy with our own practical arts and agencies. The deepest irony of all—and Kant himself is in no small part to blame for this—is that whereas Kant's single most important aim in the critical philosophy was to urge this *entirely practical replacement* for the traditional *object-oriented* rationalist metaphysics (including his own pre-Critical Leibnizian metaphysics), thus calling for a revolutionary turn to the morally practical and to the scientifically natural, it has unfortunately been what Kant's readers have taken to be *his own* indulgence in a thoroughly dubious speculative metaphysics of unknowable objects as 'things in themselves' that has continued to mar his philosophical legacy.

I have argued elsewhere that we ought to reject Sellars' quasi-transcendental idealist (and in my view, non-Kantian) contention that the object-ontology of "the common sense framework"—that is, the "manifest image" conception of perceptible, colored physical objects—"is transcendentally ideal, i.e. that there really are no such things as the objects of which it speaks" (SM V, §95; italics added; again, see O'Shea 2016b). What can and should be preserved from Sellars' account of sensory consciousness is not his specific quasi-transcendental-idealist relocational ontology (which was based on Sellars' own unquestioned but disputable qualia intuitions concerning 'ultimate homogeneity'; cf. Rosenthal 2016), but rather the more general strategy behind Sellars' then-revolutionary scientific naturalist philosophy of mind. This was the general open-ended idea that the scientific hypothesis of various representational systems in human cognition could be systematically integrated, in a non-trivial way, *within* the sort of irreducibly normative 'space of reasons' for which Sellars is now rightly famous. In different ways Paul Coates, David Rosenthal, Huw Price, Daniel Dennett, Ruth Millikan, Jay Rosenberg, Paul Churchland, and other Sellars-influenced philosophers have substantively explored ways of integrating such naturalistic theories of perceptual representation and evolved animal agency within, rather than in philosophical isolation from, various of Sellars' other distinctive views in epistemology, semantics, and metaphysics. Furthermore, it is also surely in the spirit of both Kant and Sellars, not to mention Peirce, to suggest that just which kinds of revision of our empirically manifest *object-ontology* might ultimately be required in the future, however radical, should primarily be a matter of ongoing case-by-case inquiry rather than quasi-transcendental argument. Sometimes such explanatory revisions might fit Sellars' Kuhnian-style ontological replacement model, but in other cases, particularly in biology and other special sciences, functional realization and other non-reductive models of explanatory integration are likely to remain more apt (as perhaps Wilfrid's father, Roy Wood Sellars, would have stressed).

[4] Finally, suppose now that, as just recommended, we thus reject Sellars' quasi-transcendental 'scientific noumenalism' (as it were), according to which the ordinary colored physical objects of the manifest image strictly speaking do not exist, while accepting in all other respects his otherwise Kantian and integrated scientific naturalism (again, I chart this course in O'Shea

2016b). The remaining question is whether the account of Kant's transcendental *freedom* that Sellars has offered introduces any conceptions of our thought and agency that must be inconsistent with the sort of integrated Sellarsian and Kantian naturalism that I have been attempting to delineate in this chapter. Has Kant's conception of autonomy really added anything that is somehow mysteriously beyond the evolved conceptual capacities for perception, inference, and volition that we are now assuming to be in place within a non-reductive yet fully embodied *Kantian naturalism*?

Just as we followed Sellars in trading Kant's empirical dualism for fully embodied Strawsonian personhood, however, I think that Sellars' conceptual distinction mentioned earlier between causally relevant circumstances as opposed to globally metaphysically determinist pseudo-circumstances, together with the resulting idea that we are not passive with respect to the past, would be worth exploring as taking the sting out of Kant's claim, as we saw Sellars put it earlier, that we supposedly "know, on philosophical grounds, that as objects of empirical knowledge we are *not* autonomous beings" (§86). As naturalists with that conception of *Willkür* or freedom of choice we could then plausibly agree with both of the following seemingly contradictory (but in fact merely 'Dialectically' opposed) claims by Kant in the following passage from the second *Critique*:

> One can therefore grant that if it were possible for us to have such deep insight into a human being's cast of mind, . . . that we would know every incentive to action, even the smallest, as well as all the external occasions affecting them, we could calculate a human being's conduct for the future with as much certainty as a lunar or solar eclipse and could nevertheless maintain that the human being's conduct is free.
>
> (Kant, *Critique of Practical Reason*, Ak. 5:99)

But as we have seen, Sellars in the APA address has also taken his modified agreement with Kant on freedom a step further, as follows.

As noted above, intentions and volitions for Sellars are practically efficacious thoughts that *ceteris paribus* give rise to the corresponding actions so conceived; and in that sense such intentions are intrinsically or internally motivating.[13] A genuinely autonomous *Wille* or *pure* practical reason in particular, on Sellars' reading, concerns our capacity to be motivated to action by a thought the generic or abstract *content* of which, from the moral point of view, is not specifically about any of our ends or interests in particular. Such a practically efficacious thought or 'pure idea of reason' is characterized by Kant as 'a-temporal' or 'supersensible' in part because its content is *disinterested* or non-empirical in the non-mysterious way just noted; and also because in deliberating and acting from such a motive or conceptualized point of view I must 'practically presuppose', as Kant variously puts it, that my own generic thought or idea of reason is the sufficient cause of my action.

Kant thinks that *experience* shows that we do have the capacity to act both in conformity to and in opposition to such a principle or abstractly motivating thought; and he contends that this is sufficient to *assert* our freedom without our being able to *prove* its metaphysical reality theoretically. What the critique of speculative object-metaphysics shows, according to Kant, is that the capacity for such an autonomously motivating thought or idea is not *ruled out* by anything we could ever come to know about nature from a scientific or theoretical perspective—and this is the real root of what Sellars above called Kant's "agnostic note": that at least "we cannot *know*, on philosophical grounds, that as noumena we are *not* autonomous" (I §86, quoted above).

There are of course influential ways of reading Kant's transcendental idealism that fit congenially with our having rejected Sellars' strongly 'scientific noumenalism' understood as the thesis that the objects of the manifest image strictly speaking do not exist. On such readings Kant's so-called agnosticism concerning noumena becomes functionally equivalent to restraining oneself from certain inevitably tempting but fallacy-ridden 'dogmatic' ideas, together with the realization that those particular ideas were really all along intended for practically and theoretically *regulative* uses rather than for metaphysical object-cognition. Or if one just cannot manage to see Kant himself entirely in this sweetly revolutionary practical light, then the alternative would be to leave transcendental idealism behind entirely and follow the pragmatists in pushing the regulative fallibilism of Kant's Dialectic more strongly than Kant envisioned.[14] Either way, however, a Kantian empirical realism can and should be defended as *real* realism: as a naturally embodied realism within a comprehensive normative-practical turn. And *that* Kantian naturalist vision was the central message that Sellars intended his audience to take away from his remarkable 1970 Presidential Address to the APA.[15]

Notes

1 References to Sellars' works are according to the standard abbreviations for his works listed in the References, followed by one or more of the following (depending on the work): chapter or part, sections or paragraphs (§), added paragraph numbering in recent Ridgeview editions (¶), or page numbers. I follow a standard practice among philosophers of using 'single' quotation marks for mentioning items and for 'scare quote' qualifications, reserving "double" quotation marks for direct quotations.

2 References to Kant's *Critique of Pure Reason* are to the standard 'A/B' (first edition/second edition) marginal page numbers, all translations here taken from the Guyer and Wood Cambridge edition.

3 For a sustained analysis and defense of central aspects of the views of both Kant and Sellars on the nature of the thinking self, see Jay F. Rosenberg's unjustly neglected book, *The Thinking Self* (Rosenberg 2008), recently reissued by Ridgeview Publishing Company (www.ridgeviewpublishing.com). For a basic presentation of Kant's conception of the self as it figures in the *Critique of Pure Reason*, see O'Shea (2014), especially chapter 4.

4 Sellars suggests that this pertains particularly to our sense-perceptual mental states: cf. I §§17, 20–1, 40. For the most part I will not be concerned with his various comments about "reductive materialism" insofar as they are (as the alert reader of Sellars will know) concerned with problems pertaining to qualitative sensory consciousness and Sellars' distinctive engagement with that issue.

5 For important historical and conceptual discussions of these issues, including highly knowledgeable examinations and sympathetic criticisms of Sellars' readings of Kant, within the context of a more classically metaphysical outlook on Kant's views on the soul in the first *Critique* than I am presenting here, see Ameriks (2000), e.g. pp. 62–3.

6 Even Ameriks (see 2000: 81*n*84) would in this respect appear to attribute to Sellars more than the latter is actually asserting of Kant when he (Sellars) writes that "Kant's analysis of the Paralogisms *opens the way for him to hold* . . . that the empirical I . . . is identical with (i.e., *is*) a composite physical object" (Sellars I §30, first italics added); or again when Sellars writes that Kant "*has kept the way clear for the view* that thoughts and other representations are in reality complex states of a system, and in particular, of a neurophysiological system" (Sellars MP 240, in KTM ¶59, italics added). Here "Sellars surely goes too far," according to Ameriks (Sellars MP 240, in KTM ¶59, italics added). However, in neither case, I think, does Sellars mean to suggest that *Kant* meant to "open the way" or "keep it clear" for such materialist views to be held by anyone—we've seen Sellars state that Kant denies that such materialist positions can coherently be held. Rather, Sellars is suggesting that Kant's views in this specific, highly abstract respect can be adopted and adapted *by us* in defense of such a view.

7 For a recent discussion of Sellars' views on Kant on spontaneity, with further helpful references to other recent literature on that topic (for example, Pippin 1987), see Marco Sgarbi (2012, pp. 6–9). For an excellent analysis of the overall importance to Sellars of both Kant's and post-Kantian (and in particular Hegel's) views about the nature of human reflexivity and self-determining freedom, see Terry Pinkard (2006).

8 References to 'MM' are to Kant's *The Metaphysics of Morals* (1797), using the Cambridge Edition version and the standard *Akademie* edition volume: page references.

9 For a detailed example of how this work out with respect to Kant's Second Analogy in relation to the regulative maxims of reason (in the Appendix to the Dialectic) and reflective judgment (in Kant's third *Critique*), see O'Shea (1997). One can also find in that article references to Kant commentators who find radical shifts or internal confusions in Kant's views in this domain as a result, I argue, of failing to appreciate this crucial aspect of Kant's philosophical methodology.

10 Remarkably, Valaris (2013) (section 1) refers to and argues against what he characterizes as "the so called 'relative spontaneity' view, which is championed by commentators like Wilfrid Sellars. . . ." In fact Sellars' intention is to do precisely the *opposite* of that, both as an interpretation of Kant and in his own voice. (This is not to say, of course, that there is not for Sellars an element of truth in the 'relative spontaneity' view at its own, primarily non-Kantian level, as we shall see.) Here again Sellars seems not to have done an adequate job signaling his intentions in the paper to his readers.

11 For a detailed historical analysis of Sellars' evolving views on the nature of normativity in the earlier stages of his career, see Olen (in press).

12 For a good recent examination of Kant's theoretical philosophy from a broadly Sellarsian perspective, and one that, however, similarly seeks to integrate scientific realism *within* Kant's phenomenal world rather than in competition with it, see Landy (2015).

13 Compare, in the opposite direction of fit, Sellars' view of perceptions as conceptualized thoughts that suitable subjects are normally caused to have in response to the corresponding kinds of object in appropriate situations.
14 Peirce himself provides an obvious example here. More recent Sellars-inspired pragmatists include, in different ways of course: Richard Rorty, Huw Price, Michael Williams, Robert Brandom, Robert Kraut, Carl Sachs, Steven Levine, Mark Lance, and Rebecca Kukla, among others.
15 My thanks to Deborah R. Barnbaum and David Pereplyotchik for all their help from the beginning of the *Sellars in a New Generation* conference at Kent State University in May 2015 to its fruition in this volume. My thanks also to the participants in that conference for their comments, and to Fabio Gironi for further comments.

References and Further Reading

References to Sellars' Works

FCET (1967) "Form and Content in Ethical Theory," *The Lindley Lecture for 1967* (Department of Philosophy, University of Kansas).

I (1972) "' . . . this I or he or it (the thing) which thinks . . . '," the Presidential Address to the *American Philosophical Association* (APA, Eastern Division) for 1970, *Proceedings of the American Philosophical Association* 44: 5–31. (Also reprinted in *KTM* and *ISR*). References are to section numbers.

ISR (2007) *In the Space of Reasons: Selected Essays of Wilfrid Sellars*, edited by Kevin Scharp and Robert Brandom (Cambridge, MA: Harvard University Press).

KTM (2002) *Kant's Transcendental Metaphysics: Sellars' Cassirer Lectures Notes and Other Essays*, edited by Jeffrey F. Sicha (Atascadero, CA: Ridgeview).

MGEC (1979) "More on Givenness and Explanatory Coherence," in *Justification and Knowledge*, edited by George Pappas (Dordrecht: Reidel): 169–82; reprinted in *Perceptual Knowledge*, edited by Jonathan Dancy (Oxford: Oxford University Press, 1988): 177–91.

MP (1969) "Metaphysics and the Concept of a Person," in *The Logical Way of Doing Things*, edited by Karel Lambert (New Haven, CT and London: Yale University Press): 219–52; reprinted in *KTM*.

SM (1968) *Science and Metaphysics: Variations on Kantian Themes*, The John Locke Lectures for 1965–66 (London: Routledge & Kegan Paul; reissued by Atascadero, CA: Ridgeview, 1992).

References to All other Works

Ameriks, Karl (2000) *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason*, 2nd ed. (Oxford: Clarendon Press).

Kant, Immanuel (1781/1787) *A Critique of Pure Reason*. Paul Guyer and Allen Wood (eds. and trans.) *Cambridge Edition of the Works of Immanuel Kant* (Cambridge: Cambridge University Press, 1998).

Kant, Immanuel (1788) *Critique of Practical Reason*, in Kant. *Practical Philosophy*, Mary J. Gregor (ed. and trans.) *Cambridge Edition of the Works of Immanuel Kant* (Cambridge: Cambridge University Press, 1996).

Kant, Immanuel (MM, 1797) *The Metaphysics of Morals*, in Kant, *Practical Philosophy*. Mary J. Gregor (ed. and trans.) *Cambridge Edition of the Works of Immanuel Kant* (Cambridge: Cambridge University Press, 1996).

Landy, David (2015) *Kant's Inferentialism: The Case against Hume*. Routledge Studies in Eighteenth-Century Philosophy (London: Routledge).
Olen, Peter (2016) *Wilfrid Sellars and the Foundations of Normativity* (London: Palgrave Macmillan).
O'Shea, James R. (1997) "The Needs of Understanding: Kant on Empirical Laws and Regulative Ideals," *International Journal of Philosophical Studies* 5.2: 216–54.
O'Shea, James R. (2014) *Kant's Critique of Pure Reason: An Introduction* (London: Routledge; previously Durham, UK: Acumen, 2012).
O'Shea, James R. (2015) "Kantian Reflections on the Givenness of Zahavi's Minimal Experiential Self," *International Journal of Philosophical Studies* 23.5: 619–25 (doi: 10.1080/09672559.2015.1096295).
O'Shea, James R. (2016a) "Concepts of Objects as Prescribing Laws: A Kantian and Pragmatist Line of Thought," in *Pragmatism, Kant, and Transcendental Philosophy*, edited by Robert Stern and Gabriele Gava (London: Routledge): 196–216.
O'Shea, James R. (2016b) "What to Take Away from Sellars's Kantian Naturalism," in *Sellars and His Legacy*, edited by O'Shea (Oxford: Oxford University Press): 130–48.
Pinkard, Terry (2006) "Sellars the Post-Kantian?," in *The Self-Correcting Enterprise: Essays on Wilfrid Sellars*, edited by Michael P. Wolf and Mark Norris Lance (Amsterdam and New York: Rodopi): 21–52.
Pippin, Robert B. (1987) "Kant on the Spontaneity of Mind," *Canadian Journal of Philosophy*, 449–76, reprinted in Pippin, *Idealism as Modernism: Hegelian Variations* (Cambridge: Cambridge University Press, 1997): 29–55.
Rosenberg, Jay F. (1986) *The Thinking Self* (Philadelphia, PA: Temple University Press; reissued by Ridgeview Publishing Co., 2008).
Rosenthal, David M. (2016) "Quality Spaces, Relocation, and Grain," in *Sellars and His Legacy*, edited by James R. O'Shea (Oxford: Oxford University Press): 149–85.
Sgarbi, Marco (2012) *Kant on Spontaneity* (London and New York: Continuum (now Bloomsbury)).
Valaris, Marcos (2013) "Spontaneity and Cognitive Agency: A Kantian Approach to the Contemporary Debate," *Kant Yearbook* 5.1: 107–26.

2 Toward a Sellarsian Ethics for the 21st Century

Jeremy Randel Koons

Sellars exploits a distinction between I-intentions and we-intentions in his ethical theory. While one might think that this is merely a distinction between the singular and the plural, I will argue that the distinction between these two types of intention goes much deeper than mere logical form. Rather, these two types of intention involve two different types of *reasoning*—individual reasoning and cooperative reasoning—and the latter cannot be reduced to the former. Elaborating these two types of intention in terms of these two types of reasoning provides us with an excellent tool for further developing a truly Sellarsian ethics. Next, after briefly elaborating the notion of the logic of intentions in terms of material inference, I will conclude with some comments on the categorical force of moral imperatives.

Individual Intentions vs. Group Intentions

Sellars begins virtually all of his discussions of moral theory with a discussion of intentions, and the logic of intentions. Sellars reserves the word 'shall' to express intentions (e.g., "My children shall have a good education."). Logical relations among intentions are parasitic upon logical relations among the contents of these intentions. Thus, "if 'P' implies 'Q,' then 'It shall be the case that-P' implies 'It shall be the case that-Q'" (ORAV, 86).[1]

To regiment the logic of intentions, Sellars proposes to use 'shall' as an operator, which operates on the contents of intentions. Thus:

Shall-P
P entails Q
Therefore, Shall-Q

Sellars is of the view, though, that this notion of intention does not capture the moral point of view. It is not necessarily the content of individual intentions that misses the moral point of view. One can individually intend an action that is morally right, or that is altruistic, for example. But one fails thereby to will the action *as* a moral action. To move in the direction of the moral point of view, we must move from an egocentric notion of intention

to one that better captures the idea that we are part of a group. To express such a notion, Sellars introduces the notion of a we-intention. We-intentions are expressed formally as follows:

Shall_{we} [I do A]

Two things are crucial for a we-intention to express the moral point of view. First is scope: Sellars argues that the scope of this 'we' must be "rational beings generally" (ORAV, 101). Second is the intention's content: the intention must follow from a we-intention that is categorically reasonable. For Sellars, this is the we-intention that our general welfare be maximized. (We will return to this shortly.)

I will argue that intending from the moral point of view not only involves we-intending, but also that these moral we-intentions issue from a distinct type of reasoning—cooperative reasoning—which is not reducible to or replaceable by individual reasoning. I will conclude with some comments on the categorical validity of the we-intentions that express the moral point of view, as this issue relates to Sellars' project.

Cooperative Rationality

Many discussions of rationality start with the assumption that all rationality must be conceived of as individual: individual agents are acting on intentions with contents that might at most be parallel (in Sellars' terminology), but cannot be shared. To use Sellars' example, Tom and Dick may each have an intention of the form, "It shall be the case that the war ends," but this merely reflects an "agreement in attitude," and "agreement in attitude is not an *identity of intention*" (FCET, §49). It is not clear, though, how one can get an account of moral reasoning out of individual rationality, as the ship of morality always runs aground on the rocks of free-rider problems and the like. Consider a traditional problem for utilitarianism: suppose there is a water-conservation order in effect during a drought. What would maximize utility? Plausibly, it is utility-maximizing for you to violate the order, since it produces positive utility for you, and no negative utility for anyone else. Your violation of the water order will not contribute in any appreciable way to the city's water shortage; after all, you are just one person! So an attempt to build morality on the foundations of individual rationality seems to ground morality in sand. The problem is not merely (or not even) that people are selfishly motivated—the problem arises even on a utilitarian system, which counsels impartial benevolence. The problem is more basic: it arises from asking people to reason as individuals. Each person *individually* doing what is utility-maximizing leads to an *overall* outcome that is *not* utility-maximizing.

But how else are people supposed to reason? Sellars' claim is that essential to the moral point of view is that we intend not individually, but *qua* members of the larger moral community. Thus, there is a fundamental logical

difference between intending from the individual and moral points of view. This suggests the further idea that there is also a fundamental difference between reasoning as an individual and reasoning as a moral agent—that is, reasoning as a member of a group. When I reason morally, I am not just reasoning about what I should do in a particular case. Rather, as Sellars puts it, "moral principles primarily concern the consequences of *anybody* acting in a certain way in a certain kind of circumstances" (SM, 207/7.XIII.§84).

The idea that there is a notion of team rationality, not reducible to individual rationality, is advanced by Margaret Gilbert (1989, 2001), Robert Sugden (2000), and others. We can show that there is a viable notion of cooperative rationality if we can give an example where it is clear that rationality can only be construed cooperatively, not individually. Sugden presents such an example with what he calls the Footballers' Problem. Suppose A wishes to pass the ball to B, so B can score. A can pass (and B can run) either left or right, but both must decide simultaneously, without communicating. Both can see that the chance of scoring is slightly better if both choose right instead of left (say, 11 percent vs. 10 percent). If one chooses left and the other right, the pass fails and the chance of scoring is zero. The Footballers' Problem can be presented in a traditional decision matrix:

		player B	
		left	right
player A	left	10,10	0,0
	right	0,0	11,11

Figure 2.1

What should the two players do? "The answer seems obvious: each should choose right," writes Sugden. "But paradoxically," he continues, "this obvious answer cannot be generated by the theory of individual rationality, as used in game theory" (Sugden 2000, 179). On the assumptions of traditional game theory, A will presume the rationality of B, and will go right if he thinks it is rational for B to go right. But B is in precisely the same decision situation! He will go right if he thinks it is rational for A to go right. Thus, as Sugden concludes, "We have entered an infinite regress: what it is rational for a player in a situation like A's to do depends on what it is rational for a player in a situation like A's to do" (Sugden 2000, 181).

The solution to the problem is for A and B to think not about what is rational for *each* of them to do, but what is rational for the *team* to do. And it is clearly rational for the team (meaning A and B) to go right. Again, though, this solution cannot be reached from the perspective of individual rationality; it can only be a product of team rationality. And so we have an example where rationality is irreducibly cooperative: it is a question of what

we ought to do, and cannot be reduced to separate questions about what I ought to do and what you ought to do.

One might argue that a solution to the Footballers' Problem can be reached from the perspective of individual rationality.[2] Many decision theories make available to one, before one makes a decision, credences about which action one will perform in single-player games. This can be extended to a game like the Footballers' Problem. If you are player A, a natural thing to do (since you don't know what player B will do) is give yourself a credence of .5 on either side. (Player B will, of course, do the same.) Since R,R has a greater value, collectively, than L,L, the result will be that when the above credences are taken into account, each player will choose—*individually*—to go right.

To see where the above objection goes wrong, it is best to outline the premises of the argument.[3] The argument seems to proceed as follows:

1. I don't know what the other player will do.
2. So, as a rational player, I attach a probability of 0.5 to his playing R.
3. Given this belief, rationality requires me to play R.
4. So, as a rational player, I will play R.

So far so good. Notice, though, that the argument could continue as follows, where I represent the other player's reasoning.

5. The other player doesn't know what I will do.
6. But he is rational, so he will attach a probability of 0.5 to my playing R.
7. Given this belief, rationality requires him to play R.
8. So, as a rational player, he will choose R.

One problem that becomes immediately apparent is that (8) contradicts (1), as now the argument claims both that I don't know what B will do, and that I do know (or at least can confidently predict) what B will do. Thus, instead of using individual reasoning to reason from "I don't know what the other player will do" to a conclusion about what *I* ought to do, the argument instead seems to operate as a straightforward *reductio* of the claim that I don't know what B is going to do! Thus, the argument seems to vindicate team reasoning, rather than substitute for it.

In general, Sugden writes,

> This kind of bootstrap reasoning (i.e., from the premise that I don't know what the other players will do, I infer knowledge about what he will do) can't be justified as an implication of the classical rationality and common knowledge. . .I have presented the Footballer's Problem as a problem for 'classical' game theory, which I take to assume only that the players have common knowledge of the payoffs of the game and common knowledge that each of them is rational. In this framework, it is illegitimate to appeal to ideas about what it is 'natural' for people to do. We already know that most people think it natural to play

> R—the problem is to show that this is an implication of the classical assumptions![4]

But aside from this, the very move from premise (1) to premise (2) is suspect. As Sugden writes, "this isn't a case in which one can appeal to symmetry to (try to) justify equal probabilities—the two strategies are clearly different. In general, it isn't legitimate to go from 'I don't know whether x is true or false' to 'my subjective probability for x is 0.5'."[5]

I want to suggest that there is a deeper difference than mere logical form between the two kinds of intention in Sellars' ethics. Rather, each kind of intention involves a different kind of reasoning. Elaborating these two types of intentions in terms of these two types of reasoning provides us with an excellent tool for further elaborating a truly Sellarsian ethics. Consider: individual intention (shall_{I}) is, as Sellars points out, inherently I-centered. Two people can have parallel intentions, but the logical form of their intention does not, as Sellars would say, reflect membership in a group. Individual rationality is similarly I-centered: it essentially asks what it is rational for *me* to do. And this egocentric perspective on rationality misses something crucial about the moral point of view—its inherently communal nature. Thus, it is difficult to see why it is individually rational, in the absence of external sanctions, for me to follow a water conservation order during a drought. My conservation of water doesn't benefit me. Worse, it doesn't benefit anyone else, either—again, I am just one person, so the contribution made by my cooperation here is negligible, as would be the cost of my defection. Thus, it is difficult to justify this behavior from the perspective of altruism, either.

It is this communal perspective that we-intentions were introduced to capture. When one we-intends (shall_{we}), one intends not merely as an individual, but as a member of a larger group. And of course, it is precisely this sort of group intention that should issue from an exercise of group rationality.[6] If we have a general shared intention (such as that our general welfare be promoted), it follows that it is rational for us to conserve water. Thus, it is rational—collectively rational—for each of us to adopt the following intention: I shall_{we} conserve water.

I do not have the space to argue at length for an account of cooperative rationality, but I will make a few brief comments here.[7] We can draw two related conclusions from the preceding discussion. *First*, the Footballers Problem demonstrates that cooperative rationality is not reducible to individual rationality. *Second*, both the Footballers Problem and the water conservation example demonstrate that cooperative rationality is often superior to individual rationality. There are some problems where an acceptable solution simply cannot be derived from the standpoint of individual rationality, but where cooperative rationality gives us a satisfactory analysis. This is particularly true, not surprisingly, in explaining communal practices such as morality. Thus, as in our water conservation example, we can see quite

clearly that it is rational to conserve water (that is, to adopt the intention 'I shall_{we} conserve water') for the simple reason that *our* cooperation serves *our* mutual benefit.

These two types of rationality—individual rationality and cooperative rationality—can be wedded without significant difficulty not only to Sellars' account of intention and to his notion of intentional operators, but also to his account of inferences involving intentions. But first, a brief detour: if I had a quibble with Sellars' account of inference among intentions, it would be that he treats inference among the contents of intentions as though it is just another species of ordinary logical inference, rather than what it is, viz., a special case of practical inference. Thus, when he introduces inference among intentions, he says that 'Shall-P' entails 'Shall-Q' just in case P entails Q. The plausibility of this simple account is aided by the fact that in the examples Sellars uses, the performance of Q is generally a *necessary condition* for the performance of P, and so P *really does* entail Q. (Some commentators on Sellars do the same, contributing to this difficulty in interpretation.[8]) Thus, when Sellars uses an example such as, "If Jones wants to shoot people quietly, he must (would have to) use a silencer," it is plausible that 'Shall [I shoot people quietly]' entails 'Shall [I use a silencer]', for the simple reason that using a silencer is plausibly a necessary condition on shooting people quietly.[9]

But with most inferences involving intentions, inferential relations among the contents of intentions are not so straightforward. "I eat dinner" doesn't entail "I make a quiche Lorraine," or "I make a cheese sandwich," or "I get Chinese takeout from Royal China." There is often more than one means to an end, and the multiplicity of possible means to satisfy the end means that performance of no one means is a necessary condition of the accomplishment of the end. Thus, practical inference doesn't typically involve relations of strict logical entailment; more often, it shows valid paths of practical inference available to the agent—paths from an end to satisfactory means for achieving this end. Thus, an agent might be hungry, and form the intention:

Shall [I eat dinner]

Keeping in mind that inferential relations among intentions are parasitic on inferential relations among their contents, it is clear that 'I eat dinner' in no way formally entails, for example, 'I get Chinese takeout from Royal China.' But formally valid inference was never the type that most interested Sellars. For him, *material* inference was prior in the order of explanation. Brandom (2000: ch. 2) argues that a practical inference such as

It is raining
Therefore, I should get my umbrella

is materially valid, not an enthymeme or an otherwise disguised or truncated formally valid inference. Likewise, we should regard the inference from

I eat dinner

to

I get Chinese takeout from Royal China

also as a materially valid inference, although obviously one whose validity is restricted much more to a particular set of circumstances; it is much more hedged in with *ceteris paribus* clauses and subject to defeat quite easily. It is a materially valid practical inference because it starts from an end, and infers a satisfactory means for satisfying this end.

From the fact that material inferences are not enthymemes, it does not follow that all material inference is immediate. Surely, some are; Brandom's example is. But there are also *chains* of material inference. For example, I might reason as follows: "It is raining; therefore, the streets are wet; therefore, I will take my car, rather than my motorcycle, which has poor traction when the streets are wet." In this chain of material inferences, the first inference is theoretical, and the second practical. Much of our reasoning exhibits this pattern; it is not immediate, but follows a chain of reasoning to a conclusion, with some links of the chain being theoretical, some practical. In saying that practical reasoning from an end to a means satisfying this end involves material inference, I am not committed to the claim that it involves a single inference, or only practical inferences, rather than a chain of potentially mixed inferences.

Taking into account some of these lessons about the logic of intentions, let's turn now to how our account of cooperative rationality fits with Sellars' account of practical inference involving we-intentions. Suppose, again, that we decide that the following intention is categorically valid:

It shall_{we} be the case that our welfare is maximized.

As in the case of individual intentions, practical reasoning involves reasoning from ends (i.e., the contents of particular intentions) to means of satisfying these ends. Given the end of maximizing our welfare, we can determine certain means of promoting that end. But, of course, these means are means that we must promote in concert—as a group. Thus, 'Our welfare is maximized' will materially entail, as means 'Each of us refrains from harming others,' 'Each of us keeps our promises,' etc., each with a *ceteris paribus* clause attached. Then, we conclude various derivative intentions, such as "Shall(we) be [Each of us refrains from harming others]."

It is this intermediate step—from end to means—that our account of cooperative rationality is particularly well-suited to explain. This is for (at

least) two reasons. *First*: as we noted in the case of individual reasoning, the path from end to means is not usually one of strict entailment; it is one that requires practical reasoning to find a means that will satisfy the end, given the various constraints in play. Thus, if we have an end (e.g., Prevent the reservoir from running dry), and a set of constraints (the reservoir is critically low, and we are in the middle of a drought), collective rationality can suggest an *appropriate* means of satisfying this end: each of us shall conserve water. Again, any number of different means—trucking in water, at great expense, from a distant, water-rich location; killing off half of the town's population to reduce demand for water, etc.—would achieve the end of preventing the reservoir from running dry. There is no strict entailment from end to means of achieving this end.

Second: cooperative rationality, as deployed to reason from end to means, captures an essential feature of moral reasoning that Sellars repeatedly emphasizes: "[M]embers of our community, in so far as they value the general welfare . . . [do so] not from a personal point of view—external benevolence—but *as one of us*" (SM, 221/7.XVII.§130). We have seen already how it is difficult, from the individual point of view, to reason from ends such as 'prevent the reservoir from running dry' to a means such as 'I will conserve water'. But the perspective of cooperative rationality contributes precisely what was missing—reasoning as a member of a group, to a conclusion *about what each of us ought to do*. This is precisely the content that Sellars claims a moral intention must have. Such intentions have, schematically, the categorical form, 'It shall_{we} be the case that each of us does A_i if in C_i'.[10] Thus, wedding our account of collective rationality to Sellars' account of reasoning with intentions, we can reason as follows:

1. Shall_{we} make it that [Our welfare is maximized]
2. Maximization of our welfare requires that each of us must refrain from harming others when in circumstances C_1.
3. Therefore, shall_{we} make it that [Each of us refrains from harming others, if in C_1]
4. Maximization of our welfare requires that each of us must keep our promises when in circumstances C_2.
5. Therefore, shall_{we} make it that [Each of us keeps our promises, if in C_2] Etc.

Categorical Validity

By importing means-end rationality into our account of implication between intentions, are we departing from the Kantian spirit of Sellars' project, and making all imperatives hypothetical in nature? Here, again, I think Sellars offers us a way out. Suppose we decide that the following intention is categorically reasonable:

Shall_{we} make it that [Our welfare is maximized]

Sellars' crucial move, which he makes in several of his writings (e.g., SM, Ch. 7; FCET; ORAV), is this: he argues that the reasonableness of an imperative can be *both* categorical and derivative. Intrinsic—that is, non-derivative—reasonableness is only a special case of categorical reasonableness.[11] Thus, one can reason as follows

> Shall_{we} make it that [Our welfare is maximized]
> Each of us doing A_i in C_i will (help) maximize our welfare
> Therefore, shall_{we} (I do A_i in C_i)

There are two essential features of this argument. *First*, if the first premise is an intrinsically reasonable intention, and the reasoning to the conclusion is valid, then the conclusion is a categorically (though derivatively) reasonable intention. This is the concrete payoff of Sellars' distinction between intrinsic and categorical reasonableness. Categorical imperatives are concrete articulations of what specific intentions we should have in order to satisfy the formal end of morality.

Second, the minor premise ('Each of us doing . . .') necessarily involves cooperative rationality. It involves the thought that some goal—namely, our collective welfare—is promoted by some instance of collective action.[12] This step in the process of reasoning cannot be achieved by individual reasoning, for the familiar sorts of reasons elaborated above—an individual reasoner cannot achieve collective goals; this can only be accomplished by us, thinking about what each of us should do given certain circumstances.

Combining these two observations, we can conclude that moral intentions (e.g., 'I shall_{we} conserve water') are categorically reasonable (though derivatively so), but are also intentions that can *only* be derived *via* cooperative rationality from intrinsically valid we-intentions. For, as we saw, there is no path *via* individual rationality that leads to such principles, even assuming an altruistic motivation.

So it seems clear that in principle, using means-end rationality to derive subsidiary intentions is compatible with these intentions having categorical reasonableness[13]—provided, of course, that the formal intention from which they are derived is itself intrinsically, and therefore categorically, reasonable. But now we are confronted with a question that we can postpone no longer: is there such an intrinsically reasonable formal intention?

Here, we can seek guidance from John McDowell. McDowell lays out his view in reply to Philippa Foot (1972), who argues that moral requirements are only hypothetical imperatives, since one can, without irrationality, fail to see that one has reason to act on moral imperatives. McDowell agrees that "one need not manifest irrationality in failing to see that one has reason to act as morality requires" (McDowell 1978, 13), but nevertheless thinks that there is another sense in which moral requirements are categorical imperatives. He notes that it is empty, or at best a placeholder, to say that you have reason to do A because you *should* do A. Rather, "the reason

[one has to do something] must involve some appropriate specific consideration which could in principle be cited in support of the 'should' statement" (McDowell 1978, 14). On the Kantian conception, these specific relevant considerations will be features of certain situations, as conceived by the virtuous agent. Crucially, though, we don't need to posit a desire to explain why the virtuous agent's conception of the circumstances is reason-giving. Why not? Here, we can appeal to Brandom (2000: ch. 2) for clarification. Consider the following theoretical inference:

> It is raining.
> Therefore, the streets will be wet.

Brandom urges us to consider this a materially valid inference, and argues that the impulse to see it as an enthymeme results from formalist prejudices that should have been laid to rest a century ago by Lewis Carroll. Thus, no minor premise is needed to explain why this is a good inference. But a similar move can be made in practical reasoning:

> It is raining.
> Therefore, I shall take an umbrella.

The fact that this inference is subject to defeat in no way entails that it is an enthymeme. The impulse to argue that it is only a good inference with the insertion of a minor premise to the effect that I desire to stay dry results from the same formalist prejudices that pushed us in the direction of claiming that only deductively valid theoretical inferences can be good theoretical inferences.

On McDowell's view, there is no harm in attributing a desire to someone who acts in such-and-such a way. But the desire does no independent explanatory work. The agent's conception of the situation explains why she acts as she does. Thus, the agent's understanding that it is raining explains, by itself, why she takes her umbrella before going outside. Similarly, an agent's understanding that the bully is beating up the vulnerable child on the playground is a sufficient explanation of why she intervenes; adducing a desire (say, to protect the vulnerable) adds nothing to our understanding of her action.

This Kantian picture will strike many Humeans as undermotivated. Even those who find formalism unconvincing on the theoretical side might find this analogy unconvincing. They might argue that just as a matter of human psychology, people who don't have certain desires won't output the corresponding action. Doesn't one need a desire (or an intention) to stay dry in order to get an umbrella? What if one wants to sing in the rain?[14]

What is wrong with the Humean picture is not the claim that we must appeal to conation to explain human motivation. What is wrong with the Humean picture is the thought that conation is not already involved in our

cognition of the world. On the Humean picture, we cognize a world of pure, sterile fact. Then, we react in this way or that, depending on our desires or concerns; or we 'project' value on to a neutrally-cognized, but valueless world. But on the picture I am urging here, the way we see the world is in the first place conditioned by our attitudes and concerns. A person doesn't see the crazed axe-murderer dispassionately and then, based on her desire not to be slain, run away. She sees the situation *as dangerous*; her very conception of the situation is bound up with her concerns, and this conception explains why she acts as she does.[15] It is precisely because her very conception of the situation essentially involves both cognition *and conation* that we can directly infer, "Shall [I run away]" from "There is a dangerous axe-murderer running amok."[16] And the respective contributions of these two faculties cannot be "disentangled," as McDowell (1981) puts it. The conative, motivating aspect is *already present* in the agent's conception of the situation, and adding it as a minor premise is (a) not true to the phenomenology of cases like the axe-murderer case, (b) assumes (falsely, as McDowell would argue) that such acts of conceiving can be separated into distinct cognitive and conative components, and (c) is explanatorily otiose.[17]

The Humean objection against this picture can be pressed further: while there is a clear default attitude (grounded in the near-universal survival instinct) that might be at play in the axe-murderer case, this is not true in other cases. There is no default attitude to seeing rain, or a kitchen, etc. Thus, while the anti-Humean conclusion may be plausible in some cases, it is not plausible in all.

There are at least two things to be said in response to this. First, the concerns and attitudes that inform one's conception of a situation need not be default; they can be specific to an agent, a situation, or a social role. Some might result in conceptions that are almost literally invisible to a great majority of the population. An attacker in football might see the left side defense as vulnerable and immediately make a crossing pass to his striker. A pilot might hear the stall sensor and recognizing the danger of the situation, take immediate steps to reduce the angle of attack and restore lift. A surgeon might, during an operation, see the changes in the V_5 lead of the ECG characteristic of myocardial ischemia, and (perceiving the situation as life-threatening for the patient) take immediate steps to restore adequate blood flow to the heart. All of these situations involve perceptual skills that are available only to those with particular skills; but these skills—and the particular concerns and attitudes that these individuals possess in virtue of their roles and activities—allow them to see situations in particular ways that lead to immediate, skillful action.

The second response to the Humean is more concessive. McDowell, for one, is willing to concede that in some cases, "the agent's belief about how things are combines with an independently intelligible desire to represent the action as a good thing from the agent's point of view" (1978, 22). In opposition to Brandom, McDowell suggests we read the umbrella case in

this way. However, McDowell insists that the moral (and prudential) cases should be read as above: the agent's conception of the situation by itself explains why she acts as she does. This could be because when a particular situation is conceived correctly, the particular circumstances necessarily have the reason-giving force that they do (whereas perception of the rain, or of a kitchen, need not have any *particular* reason-giving force). And indeed, McDowell contrasts a second model of the way "in which an agent's view of how things are can function in explaining his actions" (1978, 22), as distinct from the Humean model outlined above. On this second model, "a conception of how things are suffices on its own to show us the favorable light in which the action appeared" (1978, 22). McDowell suggests that prudential and moral reasons both operate in this non-Humean way.

Given the above discussion, it is therefore tempting to say moral imperatives are categorical in the following sense: their binding force is not contingent on the presence of a desire. This certainly captures an important sense in which moral imperatives are categorical, but it can only be a necessary, not a sufficient condition. For on McDowell's account, the binding force of prudential imperatives is also not contingent on the presence of a specific desire (e.g., "I desire to maintain my health"). There is, however, a more fundamental sense in which moral imperatives are categorical: if one's conception of the circumstances genuinely imposes a moral requirement, then competing considerations are not just outweighed, but silenced. Sellars doesn't state things quite so strongly (or at least not so clearly): in "On Knowing the Better and Doing the Worse" [KBDW], he at least suggests that moral requirements always override other reasons; his diagnosis of weakness of the will is that sometimes the conclusively *good* reason (the moral reason) does not end up being the conclusively powerful reason (i.e., the *motivating* reason). The implication of his discussion, of course, is that the moral reason is always the conclusively good reason. Thus, to borrow terminology from Amartya Sen's critique of traditional game theory, when one engages in cooperative reasoning and adopts a we-intention, one does not merely reason from one's preferences. Reasoning from the moral point of view involves reasoning instead from one's *commitments*, and as Sen notes, "commitment does involve, in a very real sense, counterpreferential choice" (Sen 1977, 328)—that is, choice contrary to one's *individual* preferences.

Perhaps this is unsatisfying. McDowell and Sellars conclude that moral imperatives outweigh or silence competing reasons; and I conclude that this is the sense in which they are categorical imperatives. But does this show that moral requirements are somehow mandatory, not optional? Do we have to adopt the moral point of view, to adopt the intention "I shall (we) maximize the general welfare"? McDowell, for one, regards this question as confused: "The question 'Why should I conform to the dictates of morality?' is most naturally understood as asking for an extra-moral motivation which will be gratified by virtuous behavior. So understood, the question has no answer. What may happen is that someone is brought to see things as

a virtuous person does, and so stops feeling the need to ask it" (McDowell 1978, 22). Elsewhere, McDowell argues that "the rationality of virtue . . . is not demonstrable from an external standpoint" (1979, 346).

Sellars himself seems ambivalent as to whether adoption of the moral point of view can itself be justified. Sellars ends "Science and Ethics" (and its unpublished precursor essay "'Ought' and Moral Principles") with the somewhat hopeful Aristotelian thought that "*really* intelligent and informed self-love supports the love of one's neighbor, which alone directly supports the moral point of view" (OMP, §48; SE, 200/§64). He also suggests, in the closing pages of *Science and Metaphysics*, that morality could be grounded if the following two premises could be established:

(a) To think of oneself as [a] rational being is (implicitly) to think of oneself as subject to epistemic oughts binding on rational beings generally.
(b) The intersubjective intention to promote *epistemic* welfare implies the intersubjective intention to promote welfare *sans phrase*.

(SM, 225/7.XX.§144)

However, he concedes that it would take "all the dialectical skill of a Socrates, a Hegel, or a Peirce" (SM, 226/7.XX.§144) to establish these premises, and admits that the task is beyond him. He concludes that "the argument for the reality of an *ethical* community consisting of all rational beings . . . remains incomplete" (SM, 226/7.XX.§145).

Other of his writings seem to imply that Sellars would agree with McDowell: it is not possible to justify the formal end of morality. Of particular interest are Sellars' comments (SE, OAFP) on the acceptance of the 'first principles' of a theory, from which the derivative principles of a theory are derived. Sellars casts the discussion in terms of whether "the decision to accept the theory of which these principles are the first principles can be given a rational defense" (SE, 198/§56). For Sellars, such a decision "is a piece of practical reasoning and. . .involves the relationship of means to end" (SE, 199/§57). In the case of our empirical investigation of the world (which results in the discovery of "empirical law-like statements"), "the end-in-view . . . is the state of being in a position to draw inferences concerning new cases, in a way which explains the observed cases" (OAFP, 312). What is the end-in-view of moral theory? Sellars writes,

> The only frame of mind which can provide *direct* support for moral commitment is what Josiah Royce called Loyalty, and what Christians call Love (Charity). *This is a commitment deeper than any commitment to abstract principle*. It is this commitment to the well-being of our fellow man which stands to the justification of moral principles as the purpose of acquiring the ability to explain and predict stands to the justification of scientific theories.
>
> (SE, 200/§63)

Crucially, although Sellars seems to think that theories can be 'vindicated' (his word, a usage he borrows from Herbert Feigl) via the above-described bit of practical reasoning, Sellars says little about what would, in turn, justify this end-in-view. This suggests, as I noted above, a certain ambivalence on Sellars' part about justifying the adoption of the moral viewpoint. Though he makes some (rather half-hearted) efforts to justify the formal end of morality, in other places he speaks as though the formal end of morality, this 'end-in-view' in virtue of which moral theory *tout court* is justified, cannot be given a self-standing justification or vindication. Indeed, Sellars sometimes presents the choice between the personal point of view—that of pursuing one's individual happiness through I-intentions—and the moral point of view—pursuing the collective welfare of rational humanity through we-intentions—as though it were a choice among incommensurable options. For example, in KBDW he identifies two coherent, 'all things considered' motives from which one may act:

(1) the welfare of our community viewed as related to the actions of each of us (inter-personal benevolence)
(2) one's own happiness or well-being, viewed from a personal point of view *as* one's own happiness or well-being—in traditional terms, self-love.

(KBDW, 40–1/§52)

He concludes this article with this observation about the conflict between these two motives:

> The choice is, in an important sense, between incommensurables. Which choice one makes is a revelation of what one, at that moment, *is*. It is often surprising, sometimes exhilarating, or disconcerting, even devastating—but always a revelation.
>
> (KBDW, 42/§59)[18]

All of this suggests not only some ambivalence on Sellars' part on the relation between the moral and self-interested points of view, but also some ambivalence on the question of how to ground the moral imperative.[19] Exploration of this topic will have to wait for another occasion.

Conclusion

Although the above sketch only shows some ways in which we can elaborate a modern Sellarsian ethics, I think these suggestions allow for a richer understanding of Sellars' project. But crucially, they allow us to keep Sellars firmly within the Kantian ethical tradition he claimed to be working within, and in several ways:

(1) We have explained, with help from McDowell, the way in which moral reasons are categorically valid.

(2) Using Sellars' distinction between intrinsic and categorical reasonableness, we have shown that, even though we employ means-end reasoning as a tool of moral reasoning, derivative moral principles are still categorically reasonable.
(3) With help from McDowell and Brandom, we have suggested how a Kantian philosophy of mind (motivated by a Sellarsian view of moral reasoning) can help us understand the connection between moral cognition and motivation, thereby suggesting a resolution of certain outstanding disputes in contemporary metaethics.
(4) We have tightened the connection between Sellarsian categorical imperatives and Kantian universalizability: in reasoning cooperatively, we are essentially following Sellars' dictum that "moral principles primarily concern the consequences of *anybody* acting in a certain way in a certain kind of circumstances" (SM, 207/7.XIII.§85). Thus, the moral intention "I shall_{we} conserve water" essentially follows as a consequence of reasoning—not from the individual standpoint, but cooperatively, from the intersubjective standpoint—about the consequences of each of us conserving (or not conserving) water, given the present circumstances.

Many more questions would need to be answered in a full elaboration of a contemporary Sellarsian theory. For example, how can we accommodate agent-relative duties within Sellars' framework of universalizability? How can we carve out a notion of supererogatory action? What is the precise scope of the 'we'? But brief though they are, I hope these few comments make a positive contribution toward a renewed interest in this curiously (and sadly) neglected aspect of Sellars' philosophy.[20]

Notes

1 Sellars isn't claiming that if one has the intention that P, one will as a matter of fact have the intention that Q. Rather, this must be construed as a rational principle, à la Kant's principle that one who wills the end also wills the means. See Sellars SM, 183/7.III.§18.
2 Jesse Holloway raised the following objection.
3 I am grateful to Robert Sugden for discussing this objection with me in personal correspondence.
4 Robert Sugden, personal correspondence, 21 June 2015.
5 Robert Sugden, personal correspondence, 21 June 2015.
6 Gold and Sugden (2007) argue that collective intentions *just are* those that issue from exercises of team rationality.
7 I further elaborate the notion of cooperative rationality, and defend its centrality to morality (and other normative practices) at length in Koons (2009).
8 See, for example, David Solomon (1977, 166).
9 This holds only under ordinary circumstances—that is, not in outer space, where of course all shooting is silent, or in an artificially created vacuum, etc. Even though defeaters are remote, the inference from P to Q is still a material inference, not a deductively valid inference. More about this shortly.
10 See, for example, Sellars (SM, 224/7.XX.138).

11 See, for example, Sellars (ORAV, 100).
12 How stringent does this requirement of collective action need to be? Some policies might work only under perfect, or near-perfect cooperation, while others work well with lower rates of cooperation. I discuss this issue in detail in Koons (2009: ch. 3).
13 Sellars embraces the wedding of Kantian and teleological themes in his moral philosophy. See, for example, SM, 226/7.XXI.§§146–47.
14 David Pereplyotchik pressed this objection.
15 This move has affinities with Gibson's (1979) view that we see the world in terms of *affordances* (i.e., action possibilities and things that can be used for such-and-such).
16 This move shows how Sellars' ethics addresses a problem with which Sellars was particularly concerned, namely, with how to reconcile the motivational and justificatory aspects of morality.
17 Indeed, Sellars gestured in this direction in the following passage from "Science and Ethics": "Attitudes are settled ways of viewing the world . . . [E]motivism overlooked the fact that attitudes and expressions of attitudes belong to the *rational* order. . . . It is as a rational being that man has attitudes. All commitments, scientific as well as ethical, are attitudes, *and in no case is an attitude a sensation or feeling which accompanies a 'pure' thought*." (SE, 197/§50. Second emphasis added.)
18 See also (OMP, §47).
19 DeVries (2005, 267), writes that although "Sellars cannot claim that there is an *argument* to show that the moral point of view is more reasonable than (and therefore overrides) the self-interested point of view," he nevertheless also holds that "the ability to care about others . . . is itself necessary to a full life."
20 This paper was initially presented on May 1, 2015, at the conference "Sellars in a New Generation" at Kent State University. I am grateful for the feedback I received from the audience there. In particular, Jesse Holloway, David Pereplyotchik, and David Rosenthal pressed objections that led to substantial improvements in this essay. Willem deVries, Ken Westphal, and David Pereplyotchik generously read a revised draft, and provided very helpful feedback. Finally, I am grateful to Robert Sugden for helpful correspondence regarding team reasoning.

References and Further Reading

Brandom, Robert. 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.
DeVries, Willem A. 2005. *Wilfrid Sellars*. Montreal: McGill-Queen's University Press.
Foot, Philippa. 1972. "Morality as a System of Hypothetical Imperatives." *The Philosophical Review* 81(3): 305–16.
Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
Gilbert, Margaret. 1989. *On Social Facts*. Princeton, NJ: Princeton University Press.
——— 2001. "Collective Preferences, Obligations, and Rational Choice." *Economics and Philosophy* 17: 109–19.
Gold, Natalie and Robert Sugden. 2007. "Collective Intentions and Team Agency." *The Journal of Philosophy* 104(3): 109–37.
Koons, Jeremy Randel. 2009. *Pragmatic Reasons: A Defense of Morality and Epistemology*. Basingstoke: Palgrave Macmillan.
McDowell, John. 1978. "Are Moral Requirements Hypothetical Imperatives?" *Proceedings of the Aristotelian Society*, Supplementary Volume 52: 13–29.

——— 1979. "Virtue and Reason." *The Monist* 62(3): 331–50.
——— 1981. "Non-Cognitivism and Rule-Following." In *Wittgenstein: To Follow a Rule*, edited by Steven Holtzman and Christopher Leich, 141–62. London: Routledge and Kegan Paul.
Sellars, Wilfrid. 1966 (OMP). "'Ought' and Moral Principles." Unpublished manuscript.
——— 1967a (SM). *Science and Metaphysics: Variations on Kantian Themes*. Atascadero, CA: Ridgeview Publishing Company.
——— 1967b (SE). "Science and Ethics." In *Philosophical Perspectives: Metaphysics and Epistemology*. Atascadero, CA: Ridgeview Publishing Company.
——— 1967c (FCET). "Form and Content in Ethical Theory." *The Lindley Lecture* (Department of Philosophy, University of Kansas).
——— 1974 (KBDW). "On Knowing the Better and Doing the Worse." In *Essays in Philosophy and Its History*, 27–43. Dordrecht: D. Reidel Publishing Company.
——— 1980 (ORAV). "On Reasoning about Values." *American Philosophical Quarterly* 17(2): 81–101.
——— 1988 (OAFP). "On Accepting First Principles." In *Philosophical Perspectives, Volume 2: Epistemology*, edited by James Tomberlin, 301–14. Atascardero, CA: Ridgeview Publishing Company.
Sen, Amartya. 1977. "Rational Fools." *Philosophy and Public Affairs* 6(4): 317–44.
Solomon, W. David. 1977. "Ethical Theory." In *The Synoptic Vision: Essays on the Philosophy of Wilfrid Sellars*, edited by C.F. Delaney, Michael J. Loux, Gary Gutting, and W. David Solomon, 149–88. Notre Dame, IN: University of Notre Dame Press.
Sugden, Robert. 2000. "Team Preference." *Economics and Philosophy* 16: 175–204.

Part II

Philosophy of Language and Mind

3 Pure Pragmatics and the Idea of a Metatheoretical Functionalism

Boris Brandhoff

1. Introduction

This paper is about an exegetical puzzle about the early phase of Wilfrid Sellars' philosophical work. Like many Sellars scholars, I find it helpful to think about the nine years between the publication of Sellars' first essay "Pure Pragmatics and Epistemology" (PPE)[1] in 1947 and his 1956 London lectures on the Myth of the Given as a period of continuous philosophical development in which his overarching metaphilosophical and methodological commitments settle and the outlines of his philosophical system take shape. But on the face of it, there appears to be a deep rift between Sellars' earliest three essays and the bulk of his work up to "Empiricism and the Philosophy of Mind" (EPM). At the heart of Sellars' first three essays lies a staunch commitment to a clear-cut demarcation between philosophy and empirical science. PPE, "Epistemology and the New Way of Words" (ENNW), and "Realism and the New Way of Words" (RNWW) pursue the ambitious goal of laying the conceptual and methodological foundations for a genuinely non-factualistic approach to philosophical questions—an approach that avoids the pitfall of psychologism much more carefully than does, according to Sellars, the mainstream of Analytic Philosophy in the wake of Frege, Wittgenstein, and Carnap. Apparently, the very early Sellars sees the key to success in a bold extension of the formalistic 'master strategy' of the first phase of the analytic movement to all genuinely philosophical concepts: "A philosophical concept must be decidable on purely formal grounds" (PPE, pp. 13–14; italics omitted). To live up to this claim, Sellars sets out to sketch the outlines of a *formal* metatheoretical framework that is rich enough to clarify the notion of a language *as used* in a world. In rough outline, this is the gist of Sellars' famous project of pure pragmatics.

There can be no doubt that there occurred a change of direction in Sellars' published work around the year 1948. Beginning with "Language, Rules and Behavior" (LRB), the peculiar stress on formalism and the application of formal methods in philosophy disappears. What is more, in this essay, and later in "Some Reflections on Language Games" (SRLG), Sellars even turns directly to the task of sketching the outlines of an adequate "psychology

of rule-regulated behavior" (LRB, p. 117). Did Sellars give up on the most cherished ambitions of his first three essays and succumb to factualism, after all? In this essay, I shall attempt to sketch an argument for the claim that, despite initial appearances, Sellars' post-1948 turn does not involve a significant shift in metaphilosophical commitments. In particular, I shall try to show that the special sense in which Sellars, in his earliest essays, spoke of philosophy as a "formal" enterprise remains by and large valid after 1948. Thus, there is a sense in which Sellars' philosophy remains committed to the philosophical project of pure pragmatics as envisioned in his earliest essays.

From here on, I'm going to proceed in three steps: In section 2 of this essay, I shall attempt to give a brief account of the conception of philosophy operative in PPE, ENWW, and RNWW. Our reflections on the metaphilosophical project of his earliest essays will provide the basis for a reconstruction of the aim, scope, and structure of Sellars' metatheoretical framework of pure pragmatics in section 3. In the fourth and concluding part of this paper, I'm going to defend the idea that Sellars' self-confessed formalism in his early works is based on a peculiar and highly idiosyncratic conception of the formal. To put it in a nutshell, as I see it, Sellars' commitment to a formalistic conception of philosophy comes down to a commitment to the idea that the task of philosophy is to clarify the system of functional roles that serves as a norm for our factual linguistic behavior.

2. The Metaphilosophical Project of Sellars' Earliest Essays

Sellars' earliest three essays pursue an ambitious metaphilosophical project. They constitute the attempt to arrive at an adequate account of what philosophical questions are and how they are properly addressed. In doing so, Sellars aspires to draw a clear-cut line of demarcation between philosophy and empirical science. What he is after in PPE, ENWW, and RNWW is an understanding of philosophy that does justice to the idea that the philosopher not only addresses a task different from that of the empirical scientist, but also employs distinctive conceptual resources. To this end, Sellars takes what he regards as the guiding insight of early analytic philosophy as his point of departure and sets out to extend and reinterpret it in such a way as to bring it to its full fruition. This is how he characterizes his understanding of philosophy in ENWW:

> I shall argue that philosophy is properly conceived of as the pure theory of empirically meaningful languages, and that pure semantics, as it now exists, is but a fragment of such a theory.
>
> (ENWW, p. 28)

By "empirically meaningful language," Sellars understands a language "that is used in the world it is about" (ENWW, p. 35). Any such language must

have a structure that is rich enough to enable its user to formulate "knowledge of the world in which he lives" (see RNWW, p. 53). By clarifying the concept of an empirically meaningful language, Sellars' earliest three essays, accordingly, explicate what it means to say that a language represents the world to which it belongs—as opposed to just any possible world. Against the background of such an account, Sellars goes on to "sketch a grammar" (ENWW, p. 29) for epistemological predicates such as "verified," "confirmed," or "meaningful." In this sense, the meta*philosophical* project of Sellars' earliest essays goes hand in hand with a more specific meta*epistemological* task.

Empirically meaningful languages in Sellars' sense are, of course, studied by empirical science as well. Thus, a psychologist may investigate whether the actual linguistic behavior of certain sapient organisms constitutes an awareness of their surroundings. The results of her research will contribute to a *descriptive* account of empirically meaningful languages as phenomena in the world. Such an account addresses language as "a set of socio-psychologico-historical facts" (PPE, p. 21, n. 10). Philosophers, however, are after something quite different.[2] According to Sellars, the ambition of the philosopher is to develop what he refers to as the "*pure* theory of empirically meaningful languages." In contrast to the descriptive account of linguistic phenomena provided by empirical science, philosophy is not concerned with "socio-psychologico-historical facts" about linguistic behavior, but with the norms to which it ought to conform. Of course, sociologists and anthropologists address linguistic norms as well: They attempt to *describe* norms that, as a matter fact, govern the linguistic behavior of a given group of sapient organisms. Philosophers, however, engage with linguistic norms in a different way. For Sellars, the point of a philosophical account of language is to provide a standard against which the norms governing factual linguistic behavior can be clarified and criticized (see RNWW, p. 68). Hence philosophers are free to "constitute" or "posit" systems of norms for linguistic behavior without having to assume that these norms are realized in any empirically given system of rule-regulated activity (PPE, p. 21, n. 10; see also RNWW, p. 76, n. 22). In this sense, the pure theory of empirically meaningful languages is "a priori" and "non-empirical" (RNWW, p. 53).

So far, Sellars' distinction between descriptive and pure accounts of language appears to coincide largely with Rudolf Carnap's distinction between pure and descriptive semiotics (see Carnap 1942, pp. 11–13). But Sellars places an important emphasis differently. Carnap's famous tolerance principle invites the logician to choose freely among sets of linguistic rules and explore their implications without being "hampered by the striving after 'correctness'" (Carnap 1937, p. XV; see also pp. 51–52). Sellars' approach, on the other hand, is more conservative. His aim is to make explicit "the group-grammar of epistemological predicates" (RNWW, p. 75) as they were employed by philosophers, in a way, ever since the days of Plato—and not, as Carnap put it, to venture into "the boundless ocean of unlimited

possibilities" (Carnap 1937, p. XV). As the following quotation brings out, this endeavor requires a reconstruction of an ideal system of linguistic norms to which the language we actually speak is just a crude approximation:

> Our claim is that our empirical *language* can only be (epistemologically) understood as an incoherent and fragmentary schema of an ideally coherent *language*.
>
> (RNWW, p. 78; italics in the original)

At least for the early Sellars, philosophy is not only a normative, but also, in a peculiar sense, an idealistic enterprise: It is the attempt to uncover the ideal normative standard to which we are committed as rational beings.

3. Pure Pragmatics: The Basic Idea

As the pure theory of empirically meaningful languages, philosophy faces the task of clarifying what structure an ideally coherent system of norms must have in order to constitute a language that represents the world in which it is used. As Sellars sees it, the existing metatheoretical framework of pure semantics as developed, among others, by Carnap doesn't provide the conceptual resources required for this task (see PPE, p. 5). The alternative metatheoretical framework of pure pragmatics as envisioned in Sellars' earliest three essays is supposed to provide a more comprehensive approach to the normative structure of languages. Sellars' account of pure pragmatics in PPE, ENWW, and RNWW is not only full of fascinating insights, but also notorious for its perplexing density and its terminological idiosyncrasies. For the purposes of this essay, we shall focus on the basic idea and leave many of the finer details aside.

As I see it, the most straightforward way to understand the gist of Sellars' project of pure pragmatics is to regard it as a bold attempt to enrich the conceptual toolbox of pure syntax as developed by Carnap in his 1934 book, *Logical Syntax of Language*, in surprising ways. In RNWW, Sellars characterizes pure pragmatics as the "pure theory of the application of a language, of the relation of a meaningful language to experience" (RNWW, p. 55). As a first approximation to what Sellars has in mind, we may think of an empirically meaningful language as a calculus "that is used in the world it is about" in such a way as to provide its own interpretation. From the bird's-eye view, it takes Sellars two essential steps to turn the "Procrustean bed" (PPE, p. 5) of pure syntax into a metatheoretical framework rich enough to capture the application of a language in a world: Sellars proposes, first, to regard empirically meaningful languages as a special class of calculi in Carnap's sense of the term and, second, to reconstruct the relation between language and experience as a relation between elements of a certain structure of elementary sentences formulable in such calculi as he envisions. Let us take a closer look at what these steps involve.

Following Carnap's account in *Logical Syntax of Language*, we shall think of calculi as metalinguistic rule-systems that incorporate just two types of syntactical rules (i.e., formation rules and transformation rules) (see Carnap 1937, p. 4). Formation rules define the symbols of the calculus (e.g., names and predicates) and determine the ways in which these can be combined into more complex expressions, such as sentences. Transformation rules specify the conditions under which one sentence of the calculus may be permissibly replaced by another sentence. Formation and transformation rules alone individuate symbols only with regard to their syntactical properties, and impose restrictions on the order in which they are supposed to occur (see Carnap 1937, p. 2). They do not, however, characterize these symbols, or the patterns they exhibit, as bearers of meaning (see Carnap 1937, p. 5). Hence, we may think of the symbols of a calculus as uninterpreted marks or "characters" (Carnap 1937, p. 5); all we have to presuppose is that they are elements of empirically distinguishable classes of physical occurrences—be they heaps of ink on paper, drum rhythms, or finger snaps (see Carnap 1937, p. 6). The rules of the calculus govern the abstract patterns in which these characters occur, but they do not differentiate with regard to any of their properties other than the class of symbols to which they belong and the order they exhibit. In this sense, formation and transformation rules deal with "the forms of [a] language" (Carnap 1937, p. 3) alone. Accordingly, Carnap speaks of these rules as *formal* rules, of calculi as *formal* systems of rules and of pure syntax as a *formal* discipline.

Sellars' first step toward pure pragmatics consists in regimenting the rule-structure of calculi that qualify as candidates for empirically meaningful languages. In particular, he introduces two further sets of restrictions that patterns of characters expressible in such calculi must meet. As for their formation rules, the calculi in question have a rather simple structure. Its symbols are to be laid down in the form of "explicitly listed sets of primitive relational and non-relational predicate constants" and of "an explicitly listed set of individual constants" (PPE, p. 11). Predicate constants are combined with one or more individual constants to form elementary sentences. Logical constants such as connectives or quantifiers are not provided for; hence complex sentences are not formulable. As a consequence, the calculi Sellars has in mind can do without L-rules in Carnap's sense (i.e., *logical* transformation rules) (see Carnap 1937, p. 180). They do, however, incorporate a class of what Carnap refers to as P-rules (i.e., *extra-logical* transformation rules) (Carnap 1937, p. 180). In this sense, calculi eligible for the status of empirically meaningful languages are, in Sellars' terminology, "P-lawful systems" (PPE, p. 11).[3] Throughout his earliest essays, Sellars calls the extra-logical transformation rules constitutive of P-lawful systems "conformation rules." In his subsequent work, he prefers to speak of "material rules of inference" (see, e.g., IM, p. 226). Conformation rules govern direct transitions from the statement of one elementary sentence to the statement of one or several others. Thus, the fact that elementary sentence p is

stated may necessitate with logical force that elementary sentences q and r are stated as well. Within a P-lawful system, each individual elementary sentence comes and goes as an element of a larger pattern of conformable elementary sentences.

In order to constitute an empirically meaningful language, a P-lawful system must have a distinctive structure: Its conformation rules must be set up in such a way as to individuate sets of conformable elementary sentences that are both coherent and exhaustive. Sellars calls such sets "world-stories." While the technical specifics of his proposal are a bit tricky, the guiding idea is fairly easy to grasp. Each world-story corresponds to a constellation in which, first, all elementary sentences that can be stated together at the same time are actually stated, and second, each stated elementary sentence is connected to the remainder of the set by virtue of the fact that it shares its individual constants with certain relational sentences that are also stated. Given the conformation rules of an empirically meaningful language, any world-story will be determinable by a more or less inclusive sub-set of its elements. We may think of the elements of such a subset as initial positions in the 'game' of world-story construction. Once these initial positions are taken, the rules of the calculus oblige its user to state the other elementary sentences of the world-story as well. There is no leeway for the constructor of a world-story: The calculus completes the pattern as a whole on the basis of a sufficiently large part of it.

Pure pragmatics is an attempt to answer the question of what structure an ideally coherent system of linguistic norms must have in order to constitute a language that represents the world in which it is used. We now know the first part of Sellars' answer: At the heart of an empirically meaningful language lies a P-lawful system that is sufficiently rich to allow the formulation of a world-story. Sellars thinks of such a world-story as the linguistic picture of a possible state of the world (see ENWW, p. 38; RNWW, pp. 53–54). Each of its elementary sentences designates a state of affairs. The semantic correlate of the world-story as a coherent and exhaustive set of *conformable* elementary sentences is a system of *compossible* states of affairs bound together by laws of nature (see RNWW, p. 62). As a first approximation, we may imagine an empirically meaningful language as a means for 'cognitive cartography'. Still, we should not think of the construction of a world-story as a 'use' to which such a language is put. Rather, Sellars regards the world-story formulated in an empirically meaningful language as a crucial element in the functional architecture of the language itself. Each empirically meaningful language must include a complete world-story as the "meaning-base" of the individual constants it contains (see ENWW, p. 37). Without it, its expressions would not be able to fulfill the role they are supposed to play when the language is used (i.e., to single out, and characterize, objects in the world).

Empirically meaningful languages center around a world-story; in this sense, they are committed to a unique state of the world. It is one of the guiding ideas of Sellars' project of pure pragmatics that the norms that govern the hook-up between language and world are to be specified with regard to the world the language is about. For present purposes, a cursory look at how Sellars sets out to reconstruct one central epistemological norm within the framework of pure pragmatics must suffice. In his attempt to elucidate the normative structure of empirically meaningful languages, Sellars draws on an account of verification based on the model of Moritz Schlick's notion of a *Konstatierung*. For Schlick, *Konstatierungen* are statements that refer to presently given immediate experiences (see Schlick 1934, p. 92). The truth of such a *Konstatierung* can be ascertained by directly comparing it with its referent at the very moment in which the *Konstatierung* is formed (see Schlick 1934, p. 97). Thus, *Konstatierungen* are able to bring verification processes to a definite end, and to provide an infallible, if elusive foundation of knowledge (see Schlick 1934, pp. 98–99). Just as Schlick, Sellars takes "a *verified* sentence" to be "*a sentence a token of which is coexperienced with its designatum*" (ENWW, pp. 31–32; italics in the original). However, whether such a relation holds between the token of a sentence and the immediate experience to which it refers is, for Sellars, not a fleeting fact of consciousness, but rather a question that can be unequivocally decided on the basis of the world-story of the language to which that sentence belongs. Since this world-story is supposed to be a complete story of the world in which the language in question is used, it must include a subset of sentences that constitutes what Sellars calls a "datum-biography" of the language user (i.e., a full description of his or her immediate experiences) (see ENWW, p. 31). Within such a datum-biography, we find both sentences that are about tokens of other datum-biographical sentences and sentences that are the immediate experiences designated by the former sentences. Given the complete world-story to which an empirically meaningful language is committed, it is, then, possible to read off directly which of its elementary sentences have tokens that are coexperienced with their designata and hence must count as verified.

At this point, we have to keep in mind that an empirically meaningful language is constituted by an ideally coherent system of norms. The language we actually speak, however, is but an imperfect approximation to an empirically meaningful language in the full sense of the term (see ENWW, p. 40). As such, it includes just a crude and gappy sketch of a world-story—and not nearly a complete datum-biography of the sapient organisms that use it. Sellars' account of verification is part and parcel of an attempt to clarify the ideal normative standard to which any user of an empirically meaningful language is implicitly committed. It is a most significant intermediate result that, in addition to formation rules, transformation rules, and conformation rules, this standard also includes norms that must be specified with regard to a world in which the language is used.

4. Sellars' Conception of the Formal: Metatheoretical Functionalism *Avant la Lettre*

As we have seen so far, Sellars' earliest three essays develop a vision of philosophy as an enterprise that aims to clarify languages as ideally coherent systems of norms. In his attempt to sketch the outlines of a pure pragmatics, Sellars takes on the task of providing a metatheoretical framework rich enough to apply to the normative structure of languages that represent the world in which they are used. Throughout his earliest period of work, Sellars leaves no doubt that he takes this understanding of philosophy to go along with a commitment to what appears to be metaphilosophical formalism in its most radical form. Sellars refers to philosophy as a "formal science" (PPE, p. 16) and even as "pure formalism" (PPE, p. 25), and claims that "a philosophical concept must be decidable on purely formal grounds" (PPE, pp. 13–14; italics omitted). The following quotation from RNWW brings out how close a connection Sellars sees between the normative thrust of pure pragmatics and its supposed formality:

> Pure pragmatics *or, which is the same thing, epistemology,* is a formal rather than a factual area. In addition to the concepts of pure syntactics and semantics, pure pragmatics is concerned with other concepts which are *normative* as opposed to the factual concepts of psychology, as 'true' is normative as opposed to 'believed', or 'valid' is normative as opposed to 'inferred'.
>
> (RNWW, p. 61; italics in the original)

At the present stage of our argument, it doesn't surprise us anymore that Sellars equates pure pragmatics with epistemology. For Sellars, the aim of epistemology is to clarify the norms pertaining to the cognitive dimension of language use, and it is pure pragmatics that provides the conceptual resources required to take on this task in a systematic and methodologically reflected way. Much more surprising, however, is Sellars' supposition that the idea that pure pragmatics is a formal discipline may be, to some degree, bound up with the fact that it "is concerned with other concepts which are normative." Carnap, for example, would certainly agree that the concepts explicated in pure semantics are, in a sense, normative concepts; after all, a semantical system is "a system of rules" (Carnap 1939, p. 6). Nevertheless, pure semantics is, for him, not a formal discipline, nor does he regard semantical systems as formal systems (see Carnap 1939, p. 16). We saw in the previous section that, in Carnap's understanding of the term, a *formal* approach to language is one that takes only the forms of a language (i.e., its characters as empirically distinguishable classes of uninterpreted marks and the order which they exhibit) into account. As Sellars succinctly put it in his 1954 essay "Presupposing" (PRE), an approach can count as formal in Carnap's sense if it is about "manipulating symbols without peeking

at the extralinguistic context in which these symbols are used" (PRE, pp. 198–199). In contrast to pure syntax, pure semantics does not completely abstract from "the extralinguistic context": Semantical systems incorporate rules of designation that relate the expressions of a language to objects in the world such as things or properties (see Carnap 1939, p. 9). And the same holds, of course, for Sellars' pure pragmatics. Being a special class of calculi, P-lawful systems qualify, indeed, as formal systems in Carnap's sense of the term. But as we have seen, pure pragmatics attempts to specify norms for the application of P-lawful systems as well. In doing so, it has to make reference to a world in which the calculus in question is used.

Thus, it seems reasonable to suppose that the sense in which Sellars, in the quotation above, refers to pure pragmatics as a "formal area" departs from Carnap's use of the term. Indeed, Sellars claims that the progress of analytic philosophy was impeded by "too narrow a conception of the formal" (PPE, p. 8). What is Sellars' alternative? At this point, the problem is not that Sellars had left us with too few clues about his peculiar understanding of the formal, but rather with too many. Though certainly worthwhile, a complete account of the formal/factual distinction in Sellars' early works is well beyond the scope of this section. Instead, we will focus on a single line of reasoning that is actually much more central to Sellars' overall argument than it first appears. A closer look at his reinterpretation of Peirce's type/token distinction will lead us right into the heart of Sellars' conception of the formal.

In the sentence: "A rose is a rose is a rose," we find three individual occurrences of the English word "rose." Following the terminology introduced by Peirce, we may refer to the three instances of "rose" written down above as tokens of the type *rose*. According to Peirce, tokens are single events or objects localized in space and time; types, on the other hand, are abstract entities that are embodied by these events or objects by virtue of determining one or several of their properties (see Peirce 1906, p. 506). What exactly must the three occurrences of "rose" in the sentence above embody so as to become tokens of the type *rose*? Perhaps the most straightforward answer would be to say that the property in question is an abstract sign-design. According to this interpretation, types in Peirce's sense are classes of visual or auditory patterns. In his 1950 paper "The Identity of Linguistic Expressions and the Paradox of Analysis" (ILE), Sellars introduces a handy notational device for forming expressions that refer to such classes: his star-quotes (see ILE, p. 26). By putting an expression into star-quotes, Sellars means to form a metalinguistic sortal. The extension of this sortal is the class of linguistic items that conform to the same sign-design as the star-quoted expression. Thus, the expression "*rose*" refers to the class of linguistic items that consist of an r-shaped grapheme followed by an o-shaped grapheme followed in turn by an s-shaped and an e-shaped grapheme. The intension of "*rose*," on the other hand, is the abstract sign design or pattern of marks to which the members of the class of r-o-s-e-shaped graphemes must conform. Needless to say, the *exact* scope of both

the extension and the intension of a metalinguistic sortal formed with the help of star-quotes is subject to conventions and pragmatic considerations.

According to the proposal at hand, the type *rose* is the *rose*. Tokens of the type *rose*, on the other hand, are all those linguistic items that fall into the extension of *rose*. They are members of a class of linguistic items that share the same sign-design. In his earliest essays, Sellars acknowledges the usefulness of classifying linguistic items with regard to the sign-design they exhibit. In his terminology, a star-quoted expression such as *rose* individuates a token-class. However, Sellars doesn't equate types with token-classes. For Peirce's concept of a type, Sellars has an entirely different metalinguistic use in mind. As we learn in ENWW, Sellars takes "a linguistic type" to be "a *nexus* of formal functions" (ENWW, p. 35; italics in the original). When carefully interpreted, this notion gives us a decisive hint as to what his commitment to the idea that pure pragmatics is a "formal area" actually comes down to. Throughout Sellars' earliest essays, we find passages that suggest that types in his understanding of the term pertain to languages as ideally coherent systems of norms. In the following passage from RNWW, for example, Sellars proposes to regard types as "norms or standards" that govern the use of linguistic items:

> [. . .] it will help clarify the relation of symbol types to symbol tokens, if we think of the former as norms and standards and of the latter as events that satisfy them.
>
> (RNWW, p. 52)

And in PPE, Sellars simply speaks about:

> [. . .] language as norm or type [. . .]
>
> (PPE, p. 8)

What Sellars has in mind becomes much clearer in ILE and in "Quotation Marks, Sentences and Propositions" (QMSP), another paper published in 1950. In the former essay, Sellars interprets types in his sense of the term as "linguistic roles" (ILE, p. 35). He attempts to bring out what he means by this by drawing on an analogy to the game of chess:

> If one were asked 'What is a knight in the game of chess?' one might be tempted to reply by describing the shape of a knight in one's chess set. This, however, would clearly be a mistake. [. . .] Fundamentally, then, chessmen are roles and the roles are specified by the rules of chess. Our distinction between linguistic roles, token-classes and tokens is exactly mirrored by the distinction between chess roles, chess set designs and the individual pieces in my chess set. Just as chessmen are roles specified by rules, so the roles of a language are roles specified by rules.
>
> (ILE, p. 27)

Types in Sellars' sense of the term are functional roles that are shared by linguistic items insofar as they are subject to the same, or relevantly similar, linguistic rules. The rules in question are, of course, specified with respect to token-classes. In other words, a functional role is a class of rules that govern the use of particular token-classes of linguistic items. A functional role is not—and this is a point of great importance—a second-order class of linguistic items such as the class of the token-classes that are subject to the same linguistic rules. Again, Sellars comes up with a handy notational device for forming metalinguistic sortals that pick out linguistic items according to their functional role: his dot-quotes. Thus, the intension of the expression "•rose•" is the functional role that is played, for example, by *rose*s (in English), *rosa*s (in Italian), or *gulaab*s (in Urdu). According to our previous analysis, this functional role is the class of rules that govern the use of *rose*s in English. The extension of •rose• is the class of all token-classes that are, in the language to which they belong, subject to the same rules as *rose*s are in English. The pure theory of empirically meaningful languages addresses "language as norm or type" (PPE, p. 8). According to our analysis, this comes down to the idea that its target is language as a system of functional roles constituted by an ideally coherent system of norms. Calculi are a sub-class of such systems: They are systems of functional roles constituted by formation and transformation rules alone. In Carnap's understanding of the term, only calculi are formal systems. As it now seems, the formal in the received sense is just a limiting case of the functional. Sellars' broader conception of the formal includes other modes of linguistic functioning as well. When Sellars refers to philosophy as pure formalism, he does not necessarily mean to commit himself to the application of formal methods in the received sense of the term. Instead, he champions the idea that the target of philosophical activity is language as a system of functional roles that serves as the norm for factual linguistic behavior. Without any doubt, Sellars took his divergent understanding of "formal" to be a decisive step forward. But in retrospect, Sellars' early formalism turns out to be a metatheoretical functionalism *avant la lettre*.

Notes

1 In this essay, I shall refer to Sellars' published works by using the abbreviations canonized by Pedro Amaral and Jeffrey Sicha (see Amaral and Sicha 1991). All other references follow the familiar author-year system.

2 Or rather, they ought to be. For Sellars, philosophy was, throughout its history, prone to misconstrue its insights as claims that are subject to the standard of factual truth—and hence address the same task as the claims of empirical science. In his earliest essays, Sellars refers to this aberration as "factualism" (see PPE, p. 4f.; ENWW, p. 29). Later, against the background of a more elaborate understanding of the plurality of the conceptual activities we engage in, he speaks of "descriptivism" instead (see LRB, p. 117).

3 This is clearly a variation of Carnap's term "P-language" (see Carnap 1937, p. 181).

References and Further Reading

Amaral, P. and Sicha, J.: "The Philosophical Works of Wilfrid Sellars". In: *Journal for General Philosophy of Science* 22 (1), 1991, pp. 187–193.
Carnap, R.: *Foundations of Logic and Mathematics*, Chicago: The University of Chicago Press 1939.
Carnap, R.: *Introduction to Semantics*, Cambridge: Harvard University Press 1942.
Carnap, R.: *Logical Syntax of Language*, trans. Amethe Smeaton, London: Kegan Paul Trench 1937.
Peirce, C. S.: "Prolegomena to an Apology for Pragmaticism". In: *The Monist* 16 (4), 1906, pp. 492–546.
Schlick, M.: "Über das Fundament der Erkenntnis". In: *Erkenntnis* 4, 1934, pp. 79–99.
Sellars, W. S.: "The Identity of Linguistic Expressions and the Paradox of Analysis". In: *Philosophical Studies* 1 (2), 1950, pp. 24–31. [ILE]
Sellars, W. S.: "Presupposing". In: *The Philosophical Review* 63 (2), 1954, pp. 197–215. [PRE]
Sellars, W. S.: "Some Reflections on Language Games". In: *Philosophy of Science* 21 (3), 1954, pp. 204–28. [SRLG]
Sellars, W. S.: "Empiricism and the Philosophy of Mind". In: Sellars, W. S. (ed.): *Science, Perception and Reality*, Atascadero ^{2}1991, pp. 127–196. [EPM]
Sellars, W. S.: "Epistemology and the New Way of Words". In: Sicha, J. F. (ed.): *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*, Atascadero: Ridgeview ^{2}2005, pp. 28–40. [ENWW]
Sellars, W. S.: "Language, Rules and Behavior". In: Sicha, J. F. (ed.): *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*, Atascadero: Ridgeview ^{2}2005, pp. 117–134. [LRB]
Sellars, W. S.: "Pure Pragmatics and Epistemology". In: Sicha, J. F. (ed.): *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*, Atascadero: Ridgeview ^{2}2005, pp. 4–25. [PPE]
Sellars, W. S.: "Quotation Marks, Sentences and Propositions". In: Sicha, J. F. (ed.): *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*, Atascadero: Ridgeview ^{2}2005, pp. 170–179. [QMSP]
Sellars, W. S.: "Realism and the New Way of Words". In: Sicha, J. F. (ed.): *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*, Atascadero: Ridgeview ^{2}2005, pp. 46–78. [RNWW]

4 What Jones Taught the Ryleans

Toward a Sellarsian Metaphysics of Thought

Michael R. Hicks

1. My topic is what Wilfrid Sellars called the "essentially social character of conceptual thinking,"[1] an idea he says found full flower in the philosophical tradition surrounding Wittgenstein's *Philosophical Investigations*. Recognizing the essential publicity of mental life ought not to undermine our conception of ourselves as enjoying mental lives. There *is* conceptual thinking, and it is social, public.

This optimism distinguishes Sellars from, for example, Quine, for whom anti-privacy arguments relegate talk of a mental life to "an essentially dramatic idiom" (1960, 219). Though Sellars shares Quine's dissatisfaction with a traditional ("private") conception of the mind, Sellars thinks privacy a mythical accretion, inessential to rightly conceiving the mind. This is part of what he means in characterizing his view as a "revised classical analysis" (EPM, 178).

A classical conception of the mind is, at least, "non-Rylean": it does not give way to the behaviorism Sellars finds in (Ryle, 1949). Part of Sellars' optimism is his belief that covert non-Rylean mental episodes are nevertheless public in the appropriate sense. My topic is his attempt to thread this needle, identifying covert but public episodes.

Most contemporary philosophers follow Quine to this conditional: *if* publicity were essential to the mind in the way Sellars and Wittgenstein insisted, the classical conception of the mind would have to be repudiated *tout court*, classical mental discourse being understood now as at best a dramatic idiom, a useful fiction. This conditional is available for *modus ponens* or *modus tollens*. And while many try to show that the consequent is tolerable, still more reject the antecedent and deny the essential publicity of thought.

Sellars' optimism is a resistance to the conditional as such. For Sellars, a thoroughgoing publicity is compatible with a (suitably revised) classical conception of the mind. Central to his thinking is the essential sociality of *language*. This putatively unproblematic fact is supposed to secure the sociality of classical thought. This, I intend to show, is a strategic blunder. If we help ourselves to an account of the sociality of language quite independent

of non-Rylean mental episodes, the latter will come out problematically private if recognizable at all.

2. Classical mental episodes possess, as Sellars famously put it, standing in the logical space of reasons: they are normatively related to other states.[2] They can, for instance, be appealed to in the justification of beliefs.

I take seriously the idea that an episode's standing in the space of reasons is a certain social profile, what I'll call its socio-functional role. Critical attention is often paid to the *functional* dimension of Sellars' position, which prefigures conceptual role semantics and other functionalist and holist theses in the theory of content.[3] Less attention has been paid to its sociological dimension, perhaps because Sellars often refers us to his meta-ethical writings for elaboration.[4] Nevertheless, Sellars is explicit that the functional role he envisages thought to play can be individuated only by reference to a community. "[T]here is no thinking apart from common standards of correctness and relevance, which relate what *I do* think to what *anyone ought to* think" (PSIM 16).

The social status of thought is supposed to follow from that of language. In two of his most famous presentations of this line of thought, EPM and SM, Sellars relies on what he calls the myth of "our Rylean ancestors," a group in the mists of time who had no concept of covert mental episodes.[5] Part of the point of the myth is to serve as a proof of possibility: our conception of the normative standing of covert mental episodes could have derived from that of overt episodes, verbal behavior. In EPM, the great genius Jones emerges among our Rylean ancestors, and in virtue of his teachings the Ryleans end up with our meta-conceptual repertoire, our ability to other- and self-ascribe covert mental episodes. The question that focuses my discussion is: what did Jones teach the Ryleans, in virtue of which the Ryleans ended up, well, *us*?

Sellars characterizes Jones as a scientist, a psychologist in particular, formulating a theory about his contemporaries. Covert mental episodes are his theoretical posits. Thus, if Jones is *right*, even before he disseminates his theory, the Ryleans enjoy covert mental episodes. The upshot of interacting with Jones is that the Ryleans can *notice* something they *already had*, but of which they are "totally unaware" (SM 72; see also ITM 526, quoted below). Jones, we could say, discovered the mind and taught his contemporaries to notice it.[6]

According to Sellars, "The first thing to note about the Jonesean theory is that, as built on the model of speech episodes, *it carries over to these inner episodes the applicability of semantical categories*" (EPM 187). Thus, the socio-functional standing of Jonesean thoughts is secured, so to speak, analytically.[7] This raises the question I want to consider here, for on its most natural reading, the suggestion is that it is *Jones's doing* that semantical categories apply to inner episodes. Jones, in bringing inner episodes within the cognitive ken of his contemporaries, gives them a socio-functional role.

Were that right, though, Jones could not have *discovered* normatively articulated mental episodes. At best, in virtue of his activity the Ryleans would for the first time count as engaged in such episodes. It would be only slightly misleading to say that Jones *invented* the mind. As we shall see (§4), this is not Sellars' view. But if Jones discovered episodes that were already normatively articulated, then Jones's application of semantical categories to them cannot be what *gives them* their normative status. It is tempting—though of course not mandatory—to imagine that what Jones has discovered is an innate repertoire of hitherto-unnoticed but normatively articulated episodes. Indeed, plausibly this is how Jerry Fodor (1975) argues for his famed "language of thought" hypothesis.[8]

In what follows, I take Sellars' anti-nativism for granted: though Sellars is willing to posit an innate repertoire of some sort to explain (e.g., language acquisition), he usually reserves the word "thought"—as shall I—for *conceptual* thinking, and insists that in that sense of thinking, "we learn to think in learning a language" (PSIM 32). But if Jones discovers conceptual thinking, instead of teaching us to think, it is tricky to locate anti-nativism in the account. Ultimately, by attending to this crucial point, I hope to bring into view a Sellars-inspired account that is more satisfactory, if less familiar.

3. Perhaps Sellars' clearest statement of his socio-functional account of meaning occurs at the end of LTC:

> instead of analyzing the intentionality or aboutness of verbal behavior in terms of its expressing or being used to express classically conceived thoughts or beliefs, we should recognize that this verbal behavior *is already thinking in its own right*, and its intentionality or aboutness is simply the appropriateness of classifying it in terms which relate to the linguistic behavior of the group to which one belongs.
>
> (LTC, 116–17)

The meaning of an utterance is determined by reference to the community (meanings are "public, intersubjective facts" [SE 219]), not by reference to covert states of the subject. In this sense, it is independent of thought. But Sellars' characterization of some utterances as thinkings-out-loud complicates this claim. One might have agreed that the "meaning" of an utterance is to be determined by social norms, and concluded that meaning must swing free of thought because the latter is ("of course") independent of such norms. However the social is supposed to figure in Sellars' account, it does not produce a (semantic) externalism of *this* sort. Sellars has a more ambitious goal: to use his socio-functional account of meaning to illuminate covert, non-Rylean thoughts. Meaning has a kind of conceptual priority over thought.[9]

In the myth of our Rylean ancestors, the socio-functional standing of covert thoughts is secured by the fact that Jones introduces them on analogy to overt thinkings-out-loud. As James O'Shea emphasizes, Sellars has our Rylean ancestors "characterize each other's behaviors, and in particular their verbal behaviors, in [. . .] normatively and semantically rule-governed ways" (2012, 197).[10] This rich repertoire gives Jones the analogical resources to do his work. But it also gives Sellars his problem space, for it implies the following tough constraint on his picture of the Ryleans:

> a form of linguistic behaviour be describable which, though rich enough to serve as a basis for the explicit introduction of a theoretical framework of non-Rylean episodes, does not, as thus described, presuppose any reference, however implicit, to such episodes, just as we can give an Austinian description of physical objects which is genuinely free of reference to micro-physical particles.
>
> (SM, 71)

The *circularity* objection holds that Sellars cannot meet this constraint: Jones's strategy will only suffice if his semantic vocabulary already enables at least implicit reference to covert thought. The upshot supposedly is that only private episodes can have "underived" normativity.[11]

To meet this objection, the characterization of the normativity of meaning quoted from LTC *must* not make even implicit reference to covert states. "Otherwise the supposed 'introduction' of the framework would be a sham" (SM, 71). Rather, Sellars has Jones discover thoughts as the physicist discovers micro-physical particles. The important feature of Sellars' epistemology of science is its conceptual productivity: Jonesean thoughts constitute a novel class of episodes, not already captured by Rylean categories.[12] This is what induces the problem.

For, as Rebecca Kukla (2000, 195) notes, whatever else a thought is for Sellars, it wields "epistemic authority," is "fraught with 'ought'" (TC, 212), has standing in the space of reasons. If the Ryleans did not recognize these episodes—if they needed to be discovered—then until they were discovered, there was no sociological role for them to play. Of course, for some episode to realize a socio-functional role, it need not *itself* be acknowledged; it need only be of a type that is acknowledged. Given the systematicity of language, this does not even require that a token of that type has been acknowledged: the logic of the utterance can be determined by a system of social acknowledgments. None of this, though, requires *discovery* in the way Jones is said to discover covert thinkings. This is the crucial point about conceptual novelty: Jones's analogical conception of covert thought is novel, not the sort of thing the Ryleans could have hit on without a creative scientist.[13]

If after Jones we have a conception of states that were normatively articulated prior to playing a social role, then we have a conception of normativity

that is no longer socio-functional. The alternative appears to be Kukla's conclusion, that what Jones discovered "could not, literally, have been thoughts." He could only have *made* them thoughts, invented the mind.

4. According to Kukla, Jones falsely believes himself to be a discoverer when he is in fact an inventor. He is engaged in what she calls "constitutive misrecognition," which she claims is a central feature of philosophical myth-making. Be that as it may, Sellars himself clearly means to attribute to the Ryleans covert mental lives.[14] As he says to Chisholm, "They think, but they don't know that they think" (ITM, 526). Before returning to Jones's discovery, it is worth seeing why this must be Sellars' settled view.

Kukla emphasizes that Sellars' goal is to establish an *analogy* between the epistemological status of covert thoughts and that of theoretical posits, not that covert thoughts are literal posits. (O'Shea calls it a "metamodel" [2012, 189].) Nevertheless, Sellars' fundamental analogy is between thoughts and theoretical posits. He prefaces his myth in EPM by lamenting a consequence of an otherwise heartening sociological fact: philosophy of science has come to be treated as a subfield in its own right, excusing other philosophers from taking direct notice of considerations derived from the epistemology of science. We, he implies, will need to take such notice. Moreover, he interrupts the myth at a crucial point to comment on the nature of scientific theories, drawing attention to the fact that their use of models does not undercut their realistic import.

Had Sellars thought that appreciating the mythic standing of Jones should lead us to downplay Jones's self-understanding as a discoverer, surely he would have said something. This is not decisive, but Sellars goes out of his way to prime us to read his myth as involving a discoverer.

More generally, while Kukla rightly denies that Sellars' purposes require the Ryleans to be totally coherent, the idea of Ryleans who have no covert mental episodes is more implausible than it might seem. For instance, for Sellars there is such a thing as conceptual *awareness*.[15] If the Ryleans have no covert thoughts, they enjoy no covert "acts" of awareness. I think this is untenable, but rather than focus on it, I want to consider perhaps a special case of this problem.

Kukla notes that for Sellars, authority must be *recognized* to count as authority. To recognize the authoritative standing of an utterance at least involves—for Sellars, probably consists in—*understanding* it, grasping it as the thinking-out-loud that it is. But this grasping is itself thinking. Thus, for utterances to possess normative standing, there must be understanding: the recognition in virtue of which they have that standing. Is this receptive thinking overt or covert? Ryleans can understand one another; if the Ryleans lack covert mental episodes, acts of understanding must be overt.[16] But that suggests that to understand someone the Ryleans must *talk over them*. If the Ryleans don't enjoy covert conceptual episodes at all, their communication

is going to be far more problematic than Sellars lets on. The probative value of the myth of those Ryleans is likely small.

This example displays one of Sellars' central ideas: how easy it is to be unaware of a covert mental episode. Although episodes of understanding are rarely discussed in this context, they're obviously there. This is the most important reason Jones cannot have invented the mind. *That one can have a thought without being aware of it is the revision in Sellars' "revised classical analysis."* If the suggestion that the Ryleans had thoughts of which they were unaware is what earns Jones's activity the label "misrecognition," then Sellars has not established his advertised revision.

The classical "Cartesian" analysis of covert thinking is a kind of thought-idealism: one cannot have a thought without being aware of it. Sellars' alternative is *realistic* in that it posits thoughts of which the subject (and everyone else) is unaware.[17] By contrast, the picture that follows from Jones as inventor is not a realism in this sense. There is reason to consider this interpretive option: Sellars' realism—the revision in his revised analysis—appears blatantly inconsistent with normativity consisting in a social standing. Thus the incoherence Kukla identifies runs deeper than her analysis allows. It undermines not just the minor point that Jones is a discoverer, but the major point that the essence of thought is not its perception.

This is a precarious conclusion. Given choice between his classicism and his revision, Sellars would likely abandon the classicism. In what way then does his invocation of Jones distinguish Sellars from a self-consciously anti-classical thinker like Quine?

5. The question is, how can Sellars be realistic about the normative articulation of episodes that can only be acknowledged via Jones's theoretical innovation, while also maintaining that such acknowledgment is crucial to their normative articulation? The answer to this question is often taken to lie in Sellars' anti-nativism. Before addressing that, we should consider in more detail Sellars' understanding of normative properties.

Compare two rules: "one ought to be grateful"; "one ought to say 'thank you'." In some important sense, whether you feel grateful is not *up to you*, even when a lack of gratitude rightly draws criticism. By contrast, whether you say "thank you" *is* up to you, and there are circumstances in which you ought to say "thank you" irrespective of whether you feel grateful. That you ought to say "thank you" is an "ought-to-do" (a "rule of action"); that you ought to feel grateful is an "ought-to-be" (a "rule of criticism").

For Sellars, "many rules of language *are* ought-to-dos" (thus, the rule about thanking). He insists, though, that those "of special interest to the epistemologist are ought-to-be's rather than ought-to-do's" (LTC, 97). This crucial point is supposed to secure the rule-governedness of language in general.

This distinction figures centrally in Sellars' mature work. While it does not occur explicitly in EPM, he alludes to it twice. First, at a crucial turning

point (§33 *ff*) he comments that "not all ought is ought to do, nor all correctness the correctness of *actions*." In particular, Schlick's *Konstatierungen* are not "actions" but (as he puts it later) *acts*, subject only to rules of criticism. Actions are *voluntary* in a way acts are not.

In commenting on Jones's account he returns to this theme:

> Although the theory postulates that overt discourse is the culmination of a process which begins with 'inner discourse', this should not be taken to mean that overt discourse stands to 'inner discourse' *as voluntary movements stand to intentions and motives*. True, overt linguistic events *can* be produced as means to ends. But serious errors creep into the interpretation of both language and thought if one interprets the idea that overt linguistic episodes *express* thoughts, on the model of the use of an instrument. Thus it should be noted that Jones's theory, as I have sketched it, is perfectly compatible with the idea that the ability to have thoughts is acquired in the process of acquiring overt speech and that only after overt speech is well established, can "inner speech" occur without its overt culmination.
>
> (EPM, §58.3)

This passage concludes with a comment about anti-nativism, so we are getting close to the heart of the matter. But its primary point is to articulate Sellars' non-instrumental conception of thinkings-out-loud. While there are instrumental linguistic actions, focus on them distorts the relation between language and thought.

Robert Brandom (1997) puts this point in an instructively misleading way:

> One can't think until one has learned to speak—one can't assert anything 'mentally' (think to oneself that . . .) until one has caught on to the social practice of public assertion. Thus talk is prior to thought in the order of explanation. Once one has learned simultaneously to talk and think, however, thought often precedes talk in the order of causation.[18]

Brandom's invocation of anti-nativism is overstated: all Sellars says is that anti-nativism is *consistent* with Jones's theory. This might seem just a point of emphasis, but that Brandom's reading licenses the stronger claim is important.

More immediately relevant, though, is Brandom's understanding of the limits of Sellars' anti-instrumentalism. Mature speakers sometimes have antecedent "communication intentions," so not all talk is thinking out loud.[19] Brandom emphasizes that thinkings-out-loud are *not* the result of such intentions. This *is* a core Sellarsian commitment. Where a Gricean might offer as a fundamental linguistic rule this ought-to-do:

> To convince your audience that you believe that snow is white, you ought to say "snow is white."

Sellars insists that fundamental rules are merely governed by ought-to-be's, thus perhaps:

> Asked whether snow is white, you ought to be disposed to say yes.

Such episodes, Sellars says, are displays of our "linguistic character" (SM 76).

One complication is that Sellars repudiates the communication-intention approach only as an approach to meaning, not as an approach to communication. As Gauker complains, Sellars often denies that thinking-out-loud is other-directed.[20] The suggestion seems to be that, if I am saying something *to you,* it is produced as a means to an end. But consider the rules in the last paragraph. If I answer your question spontaneously and unself-consciously, on the one hand I have said something that is essentially *to you*, but on the other hand, I have not engaged in the complex psychological process Sellars distances himself from. It is tempting to characterize the flow of a normal conversation in terms of thinking-out-loud. I will return to this.

Putting aside the question, which linguistic behaviors count as acts, let us turn to the status of linguistic acts. Brandom registers the existence of instrumental linguistic behavior by saying that "thought often precedes talk in the order of causation." This misses the point, for Sellars thinks that, in the order of being, thought is *always* prior to talk. He is at pains to deny that such a precedence relation is means-end. On Sellars' (Jones's) view, the underlying (covert) thought simply *causes* the utterance, and in that sense finds its expression, its culmination, there. On Brandom's telling, Jones taught us to reflect on speech temptations and sometimes refrain, perhaps when what we were tempted to say wouldn't fulfill our purposes. Thus Brandom's Jones is teaching us to think covertly: he is an inventor.[21] This explains how Brandom can read into this passage a declaration of anti-nativism. But Sellars only claims that Jones's theory is *consistent* with anti-nativism. So we still must ask: what role *does* the anti-nativism play?

6. Earlier I quoted Sellars' view that "one learns to think in the very process of learning to speak" (PSIM, 32). The upshot of linguistic training is not just the first language but also the capacity to think, which Jones will later discover underwrites our ability to use the first language. Rylean language-trainers are presumably only aware of the linguistic utterances, and so only attempt to produce, in Quine's phrase, outward conformity to an outward standard (1958, 5). Minimally, then, Jones's hypothesis is that in order to produce this outward conformity, Rylean teachers also need to have produced an "inward" conformity.

Of course, Quine can applaud empirical speculation about the "hardware" involved in competent language use. We ought to be such that we receive sensory stimulations, for instance, and we ought to be possessed of some complicated neural wiring. Quine can grant all this while still insisting

that Jones's "theory" that there is an analogy between a functional characterization of this wiring and the socio-functional characterization of meaning is merely a convenient recommendation, a dramatic idiom not to be taken factually.[22] Specifically, Quine suspects that the neural underpinning of intelligent behavior is more like the internal structure of a topiary than of a watch. We can *treat* one another as if our behavior is the product of shared covert states, but this is mere convenience.

This is the crux of the matter: no one—Quine included—need dispute that there are inner episodes, covert states of human language users, that can be treated as covert thinkings. What is supposed to distinguish Sellars from Quine is that Jones is a Sellarsian scientist, where the relevant dimension of Sellarsian science is its realism. This rules out an instrumentalist or fictionalist reading of Jones's theory. The states Jones posits must be *rightly* characterized as thoughts, as possessed of normative standing. Thus, such standing does not *merely* consist in being subject to ought-to-bes; it consists in being subject to ought-to-bes *of the right sort*. And a Quinean ought-to-be—speakers ought-to-be such that the dramatic idiom of psychology fits them—are not of the right sort. The only relevant sort of ought-to-be Sellars has introduced is articulated in terms of the reception of events of that type by one's community. But prior to Jones's discovery, these events are not received by the community at all. This is the problem.

7. Jones's "theory" has a puzzling feature that distinguishes it from similar explanations: he takes (some of) the explanandum to itself be the model for the explanans. An inner watch would not be a productive explanatory posit for visible watch behavior.[23] This echoes Sellars' troubling comfort with Fodorian talk of the language of thought, and thus the potential for a Sellarsian nativism. But put a different way, it might hold out hope for locating Sellars' anti-nativism.

Perhaps Jones's theory is that there are covert tokens of a (frequently) overt type. While the normative standing of the type as a whole derives from the socio-functional articulation of its overt instances, not all instances need be overt. And Jones's discovery is simply that the covert tokens are tokens of the type in question.[24]

Ryleans recognize *some* covert tokens of conceptual thinking: "thinkings-out-loud cut-short," short-circuited actualizations of dispositions to think-out-loud. But Jones's posits are supposed to go beyond what the Ryleans recognize. After all, he posits covert episodes to explain not just silent-but-intelligent activity, but also thinkings-out-loud. And the covert episodes underlying actual thinkings-out-loud aren't not-quite-actualized dispositions to think-out-loud, for by hypothesis in such cases, the disposition to think-out-loud is actualized! Per Jones, the covert episode *causes* the disposition to be actualized. We need covert episodes not in this way dependent on overt "actualizations." While relevant neurological episodes

would fit this bill, why should *they* count as instances of Rylean conceptual types?

On the current hypothesis, Jones discovers that what seemed like a conceptual truth—what qualifies as a conceptual episode must, say, be audible or visible—was a hasty generalization. The problem is that, if Jones has discovered that Rylean categories always did extend to intrinsically covert mental episodes, then Jones's discovery is that the Ryleans' language was not as austere as Sellars said it had to be. This, again, is the circularity objection.

Sellars rejects this strategy for this very reason. He claims that Chisholm's version of the circularity objection derives from a kind of "category mistake," conflating linguistic classifications with conceptual or cognitive ones.[25] Mistaking Jones's classification of covert episodes for the Rylean classification scheme that it is modeled on will lead one inevitably to the circularity objection.

Thus it is no accident that Sellars concludes that "thinkings-out-loud" are not *thinkings*—"conceptual episodes proper"—at all. It is too crude to put this by saying that thinkings-out-loud are the manifestations of thinkings. Rather, the utterances that the Ryleans would call "thinkings," and we "thinkings-out-loud," are so called because they are the manifestations of what are in fact thinkings. The framework in which we characterize spontaneous utterances as thinkings-out-loud is what Jones, qua theorist, is teaching us to abandon.[26]

Sellars' concern about circularity led him to these difficulties. To avoid them, we must not make the move that opens Sellars to the circularity objection in the first place.

Sellars' is a two-stage strategy: first, provide an account of linguistic meaning that makes *no* reference to covert episodes, and second use that account as a *model* for the covert episodes. The circularity objection held that no account adequate for the second stage could meet the constraint on the first stage. My suggestion is to abandon the two-stage strategy altogether.

This means that the Sellarsian owes an account of the nature of thought that goes beyond the analogical resources Sellars has Jones deploy, and explains how covert states can, nevertheless, be public. To fulfill this obligation, the picture of Jones as an enterprising (Sellarsian) theorist must be abandoned.

8. Consider an example. In §4 I offered utterance understandings as thinkings the Ryleans could fail to notice they engaged in silently. This served as a reminder of Sellars' "realistic" revision to the classical analysis. But this is my example, not Sellars', and so we can ask: is it illuminating to treat understandings as *theoretical posits*?

Of course understanding isn't observable in the way uttering is, nor is it a disposition to utter. But it is perfectly familiar: for instance, we train people

to understand when we train them in a language. While Sellars emphasizes the ought-to-bes that govern linguistic production, closely related ought-to-bes govern linguistic consumption. And taking linguistic consumption as paradigmatic of "covert thinking" gives immediate justice to Sellars' anti-nativism. One acquires *this* capacity to think in learning one's first language, because it is simply the passive-receptive complement to the capacity to utter.

Thus, no account of language can seriously deny the existence of covert states of understanding. I doubt Ryle's did.[27] But the problem is not just that Sellars is being unfair to Ryle, if he is. It is that Sellars *needs* the Ryle of his caricature. To avoid the circularity objection, he needs an account of linguistic behavior that makes no reference to unobservable (non-"Rylean") mental episodes. I just suggested that this is impossible.

"Our Rylean ancestors" cannot discuss semantics without the capacity to apply such vocabulary to covert thinkings. Perhaps they have not noticed such covert episodes. But it does not require *theoretical innovation* to extend the concept of the semantically determinate to the passive-receptive complement of linguistic utterance.

Thus, at least some covert thinking must, *pace* Sellars/Jones, be a token of the same type as thinking-out-loud. This is just the claim that caused Sellars to accuse Chisholm of making a category mistake. But I have made it in a way that does not suggest anything like a reverse conceptual priority, the priority of the intentionality of thought to language. This hints, then, at an alternative development of a Sellarsian metaphysics of thought.

9. Earlier (§5) I registered Gauker's complaint about Sellars' insistence that thinkings-out-loud are not other-directed. Of course, as I noted there (cf., n. 20) Sellars himself is equivocal on this point. He says it is no coincidence that conceptual thinking can be communicated (PSIM, 17), and he compares the give and take of an ordinary conversation to a "conversational dance" (NAO, 99). Thus, though he denies that linguistic acts are instrumental "actions," he is aware of their essential availability for communication. Before developing this thought, it is worth reminding ourselves why Sellars is resistant to making it central to his account.

Strawson, for instance, says that though "a man can run over his own thoughts verbally, can commune with himself in solitary speech [. . .] this must be at best a secondary employment of the medium" (2002, xv). The primary employment will always involve a *communication intention*. This latter thought is the one Sellars is most keen to deny.

To stress this, Sellars denies that solitary thinking-out-loud is secondary to its communicative counterpart. In soliloquy, our linguistic character is on display. We can accept this without following Sellars to the thought that communicative speech is secondary, needing to involve a separable communication intention. This is what led him to believe that solitary

thinking-out-loud can and must provide the conceptual home base for thinking.

If it did, the passive-receptive complement to thinking-out-loud would have its home in the secondary context of communicative speech, and so could legitimately go unnoticed. But Sellars, as we have seen, emphasizes the role of *acknowledgment* in the normative status of linguistic acts. Thus, even though he downplays the status of communication, Sellars still needs understanders, audience members who can acknowledge the authoritative significance of utterances. Thus, a Sellarsian metaphysics of thought cannot, as Sellars insinuates, consist in the categorization of *utterances* alone. Even identifying an utterance as a thinking-out-loud is locating it in a complex network of conceptual abilities essentially available to be deployed non-overtly, for instance in understanding.

The tension I uncovered in §4 came from the fact that, while Sellars' realism required that conceptual episodes can be overlooked, publicity required that conceptual episodes figure in social life. Sellars took his realism to require an analogy between overlooked mental episodes and theoretical posits. This is what stood in tension with publicity. This is why it is important that acts of understanding are not theoretical. For they are, nevertheless, easily overlooked,[28] and they are integral to appreciating the social dimension of thinking-out-loud.

In one striking passage, Sellars complains that for Kant "there is no problem concerning the cultural transmission of basic conceptual abilities." Kant has "no place for this role of the *ought-to-be*'s of language entry transitions" (Sellars, 1967, 60). The inculcation of ought-to-bes, he suggests, is the inculcation of *culture*. The apparent thought—which I at least want to endorse—is that not just our linguistic practices but also their silent conceptual underpinnings must be understood as essentially shaped by cultural setting. Conceptual thinking is essentially social. This, I have argued, requires us to abandon the meta-model of Jones as Sellarsian theorist.

10. My argument has relied on two Sellarsian premises. First, we ought to embrace publicity, giving preference to a sociological account of normativity over the idea that covert mental episodes have some underived intentionality, which perhaps social pressures cause to hook up with the conventions of public language. I have not argued for this thesis, which is becoming ever more controversial. But with Sellars, I take it that if normative properties are to constrain human activities, they must be part of shared human life from the beginning.

The second premise is Sellars' classicism, which I have put this way: there *are* covert thinkings. My example has been episodes of linguistic understanding, the intellectual process sparked by being in audible range of an utterance in one's native tongue.

These two claims suggest Sellars' thesis, that we learn to think *in* learning a language. Sellars' characterization of Jones as a discoverer forces us to take his words here perfectly seriously: "linguistic" training is an inculcation of a series of norms governing not just what sounds one ought to make but, for instance, how one ought to respond to sounds others make and how one ought to navigate one's environment in the absence of sound. All of this is thinking.

In appealing only to observable processes like responding to others and navigating one's environment, I display sympathy for Sellars' methodological behaviorism. But taking this last example, the ought-to-bes governing the navigation of the environment are conceptually rich, presupposing the capacity for silent sensitivity to the layout of one's environment. Plausibly, the teachers need not be self-conscious about this unity: we don't think of learning to sit at the table, or learning to walk, as of a piece with learning to talk.

So while no Sellarsian theorist, Jones can discover that intelligent behavior is conceptually structured behavior. He can unify such conceptually structured behavior with the most obvious conceptually structured behavior, speech. But departing explicitly from Sellars' presentation of him, Jones need not posit that standing behind all intelligent activity is a suite of intrinsically covert abilities somehow comparable to speech. Rightly conceiving of *speech* requires recognizing that the abilities actualized there can already be actualized silently. What we need—and I concede I have not provided it here—is an account of just what these abilities, and their actualizations, are. We need, that is to say, a metaphysics of thought.[29]

11. If I am right, there are prospects for an account that fulfills our fundamental Sellarsian desiderata: it is both anti-nativist and realist about conceptual abilities, while maintaining the essential publicity of normativity. The suggestion is somewhat ironic. Sellars motivated the publicity of thought by reflection on the analogy between language and thought, taking it, first, that the publicity of meaning was plausible in its own right, and second, that the analogy between language publicly construed and thought could underwrite a non-mysterious conception of covert mental episodes. But we have seen that this analogy can be overplayed, in a way that could have contributed to widespread relapse into mental privacy.

Taking language to be explicable *without reference to thought* opens us to the confusions of Sellars' picture of our Rylean ancestors. To conceive language as a social phenomenon requires the recognition of covert as well as overt mental episodes. One cannot, therefore, engage in the theory of meaning in abstraction from the philosophy of mind.

Notes

1 (PSIM, 1962a, 16).
2 "In characterizing an episode or a state as that of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says" (EPM §36, 169). As John McDowell (1998, §3) emphasizes, *knowledge* is merely an interesting instance of such a standing.
3 Sellars's proto-functionalism about *meaning* is not his proto-functionalism about mental states. The latter, though disputable (and disputably Sellarsian: compare Rosenberg [2004] and O'Shea [2012]), borders on being a commonplace. The former is controversial at best.
4 Cf., NAO, 110 n. 8; SM 75, §37.
5 NB: "One serious but common mistake with regard to Sellars' myth of Jones [. . .] is to think that he has presented his pre-Jonesean Ryleans in such a way that they are supposed *not to have the concepts of thoughts* . . ." (O'Shea, 2007, 102). They merely lack the concept of *covert* thoughts.
6 Late in his career, Sellars suggested that "perhaps the most basic" form of the myth of the Given was the principle O'Shea calls the myth of the categorial given: "If a person is directly aware of an item which has categorial status C, then the person is aware of it *as* having categorial status C" (FMPP, 11). This might suggest that Jones's discovery is a discovery of the fact that such-and-such (familiar) events are thoughts. I need take no stand on this now. Regardless, the Ryleans were unaware of covert thoughts *as thoughts*. Until Jones came along, the Ryleans lacked (as we could put it) categorial awareness of thoughts.
7 Compare his correspondence with Rosenthal, p. 501, §17, published as a "supplement" to Marras (1972).
8 Christopher Gauker reports a conversation in which Sellars said "that his main difference with Fodor concerned the innateness of the language of thought. Apart from that, he seemed to be saying, Fodor's idea was basically all right" (Gauker, 2011, 135 n. 10). Around the same time, Sellars wrote: "There is nothing self-contradictory about the idea of an innate language-like structure (Mentalese), and there might just be no other way of explaining language acquisition" (BLM, 5). To my knowledge he never succumbs to the suggestion that this innate mental repertoire is conceptual.
9 Compare also Michael Dummett's claim that judgment "is the interiorization of the external act of assertion" (1973, 362). We might imagine the linguistic community forming assertional practices, which can then be internalized as covert judgments. If Jones were the agent of this internalization, this would have him inventing, rather than discovering, covert mental episodes. Sellars' view isn't even an externalism of this sort.
10 Moreover, O'Shea comments, the Rylean picture prior to Jones is "nearly an entirely adequate account of our common sense or 'folk' psychology" (200). It is a mistake to see Jones as positing folk psychology; he is, at best, complicating it.
11 The locus classicus for this concern is the correspondence with Chisholm (ITM). O'Shea (2007, ch. 4) has detailed discussion and citations.
12 Sellars claims the "irenic instrumentalism" of Ernest Nagel presupposes that theoretical significance depends on predicates being "given a place [. . .] within the antecedent observation framework." This would "rule out the idea that theory construction could be a technique for enriching or revising this framework" (SRI, 170). O'Shea (2012, 200) makes clear that this kind of thinking is at stake in the myth of Jones.
13 Nevertheless, the idea that covert thinkings just are tokens of Rylean types, in this extended sense, is an attractive one, to which we shall return (§7 below).

14 Nothing in what follows counts against Kukla's understanding of philosophical myth. Both it and her discussion of Sellars deserve greater attention than I can give here. Especially illuminating is her observation that the myth of the Given's inadequacy cannot derive *simply* from its mythic status: the myth of the Rylean ancestors, though a myth, is supposed to work (it is a "myth to kill a myth"). The myth of the Given, qua myth, fails.

15 This is what he means when he somewhat notoriously says "all awareness of abstract entities—indeed, all awareness even of particulars—is a linguistic affair" (EPM 160) but compare (MEV, 325). See DeVries (2005, 185) for discussion.

16 Invoking the *ability* to understand will not help, for what we need is an *actualization* of that ability. Note also that there need be no fact of the matter as to whether the Rylean audience member has "rightly" understood, so long as her passive response to the utterance gives her *reasons* to interact with the speaker in various ways.

17 The suggestion of n. 6 above seems to back down from Sellars's realism.

18 For purposes of comparison, I quote the entirety of both Sellars's §58.3 and Brandom's comment on that subsection.

19 "Statings"—Gricean assertions—are paradigmatic linguistic actions (as opposed to acts). Cf., NAO 66, SM 75, LTC 111 (the example there is "telling").

20 Cf., SM 157: "what we are concerned with is those verbal episodes which occur when our citizens are by themselves, or when, as it were, they are overheard . . ." Gauker characterizes such events as "utterly gratuitous epiphenomena" (Gauker, 2011, 126) inadequate to Jones's explanatory purposes. But contrast NAO 98–9, §9. Notably, although in MFC Sellars had written that linguistic acts "must not be construed as other-directed social actions" (420), he left this sentence out of the reprint of that essay in NAO. (Cf., 66 for the corresponding discussion, where instead he refers the reader forward, presumably to the vicinity of p. 99).

21 Compare Brandom's Jones to Dummett (n. 9 above). Elsewhere in his study guide, I should note, Brandom is admirably clear that Jones is a discoverer. This passage, though suggestive, is aberrant.

22 Ruth Millikan (e.g., 1986) argues that biology affords an explanation of psychological idiom not—contra Sellars—"fraught with ought." (She explicitly disavows the normativity of norms at [Millikan, 2005, 83].) Millikan notes the tension between conceiving Jones as a scientist and claiming that psychological characterizations locate subjects in a different logical space than the space of scientific explanations. Millikan and Quine agree in endorsing Sellars' claim that "in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not" (EPM §41, 173); and in disputing his "classical" commitment to thoughts located in a different logical space, the space of reasons. Quine accepts Sellars' claim that thoughts would have to be located in a different logical space, and concludes that there are no such things (the attribution of them is a dramatic idiom). Millikan repudiates logical distinctness in favor of the more tractable distinction between biological and physical explanation. Both reject Sellars' classicism.

23 The curiosity here stems from the apparent explanatory status of the theory. Molecular theory, which Sellars tells us involves an analogy to billiard balls, does apply to billiard balls, but none of the relevant properties of billiard balls (i.e., how collisions work) is explained by this analogy. By contrast, the relevant intelligent properties of linguistic utterances are, somehow, supposed to be explained by positing language-like covert states.

24 I thank Jaroslav Peregrin, Christopher Gauker, and Paul Roth for independent discussions of this point.

25 SM 160 *ff*. (The reference to Chisholm is on 162.)
26 Just as we cannot *really* do without the common sense world of tables and chairs, we also cannot do without the framework of thinkings-out-loud, conceived as the Ryleans do. There are complicated questions here about what deVries (2005, 271, *ff*.) calls "practical reality" (see also deVries, this volume). Sellars finds it important to stress—as in the passage from LTC—that when we are conceiving thinkings-out-loud as the Ryleans do, we are not conceiving utterances as the culmination of inner episodes. The category mistake is to imagine that when we have thinkings-out-loud in view as thinkings in their own right, we can also see them as standing in relations to inner episodes. This precludes construing the relation between Rylean and non-Rylean episodes as either "causal" or "inferential." What non-Rylean episodes do is cause utterances, which can, from a different vantage, be seen as Rylean episodes. Because the latter are the model for the former, they engage in closely related inferential patterns. But it is a category mistake to imagine that they bear inferential relations to one another.
27 Ryle spent a good portion of the later part of his career thinking about Rodin's famous "Thinker," reflecting, enjoying mental episodes, but not displaying any changes in behavior or behavioral disposition. See the papers collected in (Ryle, 1979). Ryle might never have come to terms with the problem of covert mental episodes, but it isn't clear that an account of them must be fundamentally non-Rylean.
28 Note also that, to the extent that they are recognized at all, acts of understanding are recognized as thoughts. Thus, I take it I am not backing down from Sellars' realism in the manner suggested in n. 6 above.
29 Just what this explanatory obligation consists in is itself a matter for philosophical dispute. Though I mean to be agnostic here, I would not require that such an account be, as McDowell (1994, e.g., p. 34) put it, from "sideways on."

References and Further Reading

Brandom, Robert (1997): "Study Guide," in Wilfrid Sellars, *Empiricism and the Philosophy of Mind*. Cambridge, MA: Harvard University Press, 97–122.
deVries, Willem A. (2005): *Wilfrid Sellars*. Montreal & Kingston: McGill-Queen's University Press.
Dummett, Michael (1973): *Frege: Philosophy of Language*. Cambridge, MA: Harvard University Press, 2nd Edition, 1981.
Fodor, Jerry (1975): *The Language of Thought*. Cambridge, MA: Harvard University Press.
Gauker, Christopher (2011): *Words and Images*. Oxford: Oxford University Press.
Kukla, Rebecca (2000): "Myth, Memory and Misrecognition in Sellars' 'Empiricism and the Philosophy of Mind'," *Philosophical Studies*, 101, 161–211.
Marras, Antonio (ed.) (1972): *Intentionality, Mind, and Language*. Urbana, IL: University of Illinois Press.
McDowell, John (1994): *Mind and World*. Cambridge, MA: Harvard University Press.
——— (1998): "Sellars on Perceptual Experience," in *Having the World in View*. Cambridge, MA: Harvard University Press, 2009, 3–22.
Millikan, Ruth Garrett (1986): "Thoughts without Laws; Cognitive Science with Content," *The Philosophical Review*, 95, 47–80.
——— (2005): "The Son and the Daughter: On Sellars, Brandom, and Millikan," in *Language: A Biological Model*. Oxford: Clarendon Press, 77–91.
O'Shea, James (2007): *Wilfrid Sellars: Naturalism with a Normative Turn*. Cambridge: Polity Press.

——— (2012): "The 'Theory Theory' of Mind and the Aims of Sellars' Original Myth of Jones," *Phenomenology and the Cognitive Sciences*, 11, 175–204.
Quine, W.V. (1958): "Speaking of Objects," in *Ontological Relativity and Other Essays*. New York: Columbia University Press, 1969, 1–25.
——— (1960): *Word and Object*. Cambridge, MA: MIT Press.
Rosenberg, Jay F. (2004): "Ryleans and Outlookers: Wilfrid Sellars on 'Mental States'," *Midwest Studies in Philosophy*, 28, 239–265.
Ryle, Gilbert (1949): *The Concept of Mind*. New York: Barnes & Noble.
——— (1979): *On Thinking*. Oxford: Basil Blackwell.
Sellars, Wilfrid (1956): (EPM) "Empiricism and the Philosophy of Mind," in Wilfrid Sellers, Science, *Perception and Reality*. London: Routledge and Kegan Paul, 1963, 127–196.
——— (1957): (ITM) "Intentionality and the Mental," *Minnesota Studies in the Philosophy of Science*, 2, 507–539, a symposium by correspondence with Roderick Chisholm.
——— (1960): (SE) "Science and Ethics," in Wilfrid Sellers, *Philosophical Perspectives: History of Philosophy*. Atascadero, CA: Ridgeview Publishing, 1977, 209–232.
——— (1962a): (PSIM) "Philosophy and the Scientific Image of Man," in Wilfrid Sellers, *Science, Perception and Reality*. London: Routledge and Kegan Paul, 1963, 1–40.
——— (1962b): (TC) "Truth and 'Correspondence'," in Wilfrid Sellers, *Science, Perception and Reality*. London: Routledge and Kegan Paul, 1963, 197–224.
——— (1963): (SPR) *Science, Perception and Reality*. Atascadero, CA: Ridgeview.
——— (1965a): (SM) *Science and Metaphysics: Variations on Kantian Themes*. Atascadero, CA: Ridgeview.
——— (1965b): (SRI) "Scientific Realism or Irenic Instrumentalism: A Critique of Nagel and Feyerabend on Theoretical Explanation," in Wilfrid Sellers, *Philosophical Perspectives: Metaphysics and Epistemology*. Atascadero, CA: Ridgeview Publishing, 1977, 157–189.
——— (1967): (KTE) "Some Remarks on Kant's Theory of Experience," in Wilfrid Sellers, *Essays in Philosophy and its History*. Dordrecht: D. Reidel, 1974, 44–61.
——— (1969): (LTC) "Language as Thought and as Communication," in Wilfrid Sellers, *Essays in Philosophy and its History*. Dordrecht: D. Reidel, 1974, 93–117.
——— (1974a): (EPH) *Essays in Philosophy and Its History*. Dordrecht, Holland: Reidel.
——— (1974b): (MFC) "Meaning as Functional Classification (A Perspective on the Relation of Syntax to Semantics)," *Synthese*, 27, 417–437.
——— (1977): (PPME) *Philosophical Perspectives: Metaphysics and Epistemology.* Atascadero, CA: Ridgeview.
——— (1979): (NAO) *Naturalism and Ontology*. Atascadero, CA: Ridgeview.
——— (1980): (BLM) "Behaviorism, Language and Meaning," *Pacific Philosophical Quarterly*, 61, 3–25.
——— (1981a): (FMPP) "Foundations for a Metaphysics of Pure Process: The Carus Lectures," *The Monist*, 64, 3–90.
——— (1981b): (MEV) "Mental Events," *Philosophical Studies*, 39, 325–345.
Strawson, P.F. (2002): *Logico-Linguistic Papers*. Aldershot: Ashgate, 2nd Edition.

5 Sellars and Psycholinguistics

David Pereplyotchik

§1. Introduction

Wilfrid Sellars had an abiding interest in science, language, and perception, and his followers share his fascination with these topics. It's surprising, then, that relatively little work has been done to connect Sellars' philosophy to the science of language perception—psycholinguistics. In what follows, I bring Sellars' views to bear on live issues in this field. My hope is that this discussion opens the door to a fruitful dialogue between philosophers of a broadly Sellarsian persuasion and theorists whose research pertains more directly to psycholinguistics. To get the conversation started, I will highlight a dozen or so aspects of Sellars' overall theoretical framework, and draw on them to shed light on what's at stake in debates over the "psychological reality of grammar." The cluster of issues that fall under this heading can all benefit from the clarification that Sellars' conceptual toolkit allows us to effect.

Sellars espoused a realist attitude toward theoretical posits, and he held that psychological states are akin to the posits of science, so he was a realist about them too (EPM).[1] Applied to the domain of psycholinguistics, his view entails that we should take seriously the idea that language comprehension is causally mediated by whatever internal states the psycholinguist posits in accounting for the available behavioral and neurocognitive data. In §2, I survey several experimental results that speak in favor of positing what I'll call "mental phrase markers"—subpersonal representations of the syntactic structure of incoming linguistic input. I then explore how a Sellarsian should think of such states by asking, first, how they differ from the familiar posits of the genius Jones (EPM), and second, how to characterize their representational properties.

In §3, I argue that Sellars has the resources to both distinguish between personal-level and subpersonal representations, and to clarify the relations between them. Turning to the question of representation, I appeal to Sellars' view of meaning as functional classification (MFC) and apply his functional-role semantics (IM, SRLG) to the special case of mental phrase markers. Sellars explicitly recognized the flexibility of functional classification, which I contend can be profitably exploited in giving an account of the representational dimension of subpersonal states. Although it is clear that

such states differ significantly from personal-level thoughts and intentions, I argue that we can make sense of their representational properties by exploiting an analogy to personal-level functional roles.

Finally, in §4 I consider a different kind of psycholinguistic posit—what I'll call "mental syntactic rules," which comprise a mental *grammar* of a natural language.[2] Many psycholinguists follow Noam Chomsky in claiming that such rules are "psychologically real" in that they underlie a competent speaker's linguistic abilities. But there is reason to doubt that such states are psychologically real in the very same sense that mental phrase markers are. To clarify the debate surrounding this issue, I'll develop an analogy to Sellars' distinction between ordinary and auxiliary positions in a language game (SRLG). Sellars held that auxiliary positions can be dispensed with at the cost of multiplying moves in the game. I suggest that the mental syntactic rules are analogous to auxiliary positions in this regard. To put it in terms familiar from MEV, the human sentence processing mechanism may well be a (subpersonal) Humean system, not an Aristotelian one. I end by discussing the consequences of this view, drawing on still another distinction from Sellars (LTC), between psychological dispositions and the mental events that are their manifestations.

§2. Some Aspects of Contemporary Psycholinguistics

Psycholinguistics divides into studies of language acquisition and language use. Acquisition has received the lion's share of attention in philosophical discussions, owing largely to the fascination that philosophers have with the innateness hypothesis that Chomsky (1965) famously advanced. The present discussion bucks that trend, setting aside the vexed innateness controversy and focusing exclusively on what has been learned in recent decades about language comprehension in competent adults.[3] Similarly, whereas philosophers tend to focus on the semantic properties of linguistic items, I will be concerned here primarily with their syntactic properties. Semantics will come into the picture when we turn to Sellars' account of representation.

For many psycholinguists, the starting point for investigating language comprehension is the assumption that the human mind contains a language-processing mechanism whose function is to generate a class of subpersonal states that represent the language-specific properties of incoming linguistic stimuli. Here is a typical statement of this assumption:

> Most models of human language comprehension assume that the processor incorporates words into a grammatical analysis as soon as they are encountered. . . . We assume that sentence processing involves the computation of dependencies between the words and phrases that are encountered.
>
> (Sturt, Pickering, Scheepers, and Crocker, 2001: p. 283)

I call such representations "mental phrase markers" (MPMs), thinking of them as internal functional states whose representational content is like that of the metalinguistic description in (1).

(1) [$_{S}$ [$_{NP}$ [$_{Pronoun}$ **I**]] [$_{VP}$ [$_{Verb}$ **prefer**] [$_{NP}$ [$_{Det}$ **a**] [$_{Nom}$ [$_{N}$ **morning**] [$_{NOM}$ [$_{N}$ **flight**]]]]]]

As noted above, many psycholinguists also claim that the comprehension mechanism houses the syntactic rules constitutive of the hearer's language—rules that, *inter alia*, determine the syntactic categories of the simple units of the language and the syntactic structures of the complex units that are built from them. A leading figure in the field, Janet Dean Fodor, expresses the view as follows: "[L]et us suppose (as we surely should, until or unless the facts dictate against it) that the human sentence processing routines compute for a sentence the very structure that is assigned to it by the mental 'competence' grammar" (1989: 157). As an example, consider the following toy grammar (adapted from Jurafsky and Martin, 2008), consisting of lexical rules and structure rules. This grammar is capable of generating the string 'I prefer a morning flight' and assigning to it the structure that is rendered explicit in (1).[4]

I use the label "mental syntactic rules" (MSR) to refer to elements of the hypothesized mental grammar posited by Chomskyan linguists—a set of interacting internal states whose representational contents are relevantly like (fn. 4) the rule $S \rightarrow NP\ VP$. I will return in §4 to the psychological reality of MSRs. Here, I survey three lines of behavioral and

Lexical rules

Noun → flight | trip | morning
Verb → is | prefer | want | fly | leave
Adjective → cheapest | non-stop | direct
Pronoun → me | I | you | it
Determiner → the | a | this
Preposition → from | to | in
Complementizer → that

Structure rules	**Examples**
NP → Pronoun	I
NP → Det Nominal	a + flight
Nominal → Nominal Noun	morning + flight
Nominal → Noun	flight
VP → Verb NP	want + a flight
VP → Verb NP PP	leave + Boston + in the morning
VP → Verb Complementizer S	prefer + that + the flight leave this morning
S → NP VP	I + want a morning flight

Figure 5.1 A toy context-free grammar, adapted from Jurafsky and Martin (2008).

neurocognitive evidence in favor of positing MPMs in our account of language comprehension.

(i) Competent adults are fast at noticing ungrammaticality. Various types of syntactic violation have been found to generate distinctive neural signatures, giving researchers a handle on the timing, location, and neural underpinnings of MPMs (Bornkessel-Schlesewsky and Schlesewsky, 2009). Phillips and Lewis (2013) point out that "if the on-line analyzer is able to immediately detect any grammatical anomaly that it encounters, then it is reasonable to assume that it is constructing representations that include sufficient grammatical detail to detect those anomalies" (20).

(ii) A much-exploited phenomenon in cognitive psychology known as *priming* occurs when a mental representation is activated by a stimulus for one task but continues to influence cognitive operations on later tasks. It has been found that syntactic structures can serve as primes. In a typical study of syntactic priming, subjects are initially presented with a sentence that has a particular structure, and then asked to describe an event, presented as a picture or video, that bears no obvious relation to the sentence they were shown. Though the pictured event can be described using either of two semantically equivalent forms (e.g., [*x* gave *y* to *z*] or [*x* gave *z y*]) subjects tend to employ the structure of the sentence with which they were "primed." Pickering and Ferreira (2008) stress the importance of priming data to our understanding of sentence processing:

> In the past couple of decades, research in the language sciences has revealed a new and striking form of repetition that we here call *structural priming*. When people talk or write, they tend to repeat the underlying basic structures that they recently produced or experienced others produce. This phenomenon has been the subject of heavy empirical scrutiny. Some of this scrutiny has been because, as in other domains in cognitive psychology (e.g., priming in the word-recognition literature), the tendency to be affected by the repetition of aspects of knowledge can be used to diagnose the nature of that knowledge. . . . [T]he tendency to repeat aspects of sentence structure helps researchers identify some of the representations that people construct when producing or comprehending language. . . . [M]uch structural priming is unusually abstract, evidently reflecting the repetition of representations that are independent of meaning and sound. This is therefore informative about how people represent and use abstract structure that is not directly grounded in [phonological] or conceptual knowledge. One possibility is that the representations that it identifies can be equated with the representations assumed in formal linguistics.
>
> (p. 427)

(iii) Perhaps the most persuasive case for the reality of MPMs comes studies of "garden-path" sentences, like (2)-(7), which are fully grammatical, but often present problems, even for proficient readers.

(2) The horse raced past the barn fell.
(3) The cotton clothing is made of grows in Georgia.
(4) We all knew the man who believed the Queen hurt himself.
(5) The soldier convinced the radical student that he was fighting in the war for to enlist.
(6) The plumber told the clients he was having trouble with that he was sick.
(7) Fat people eat accumulates.

A garden-path effect occurs when the parsing mechanism selects the wrong analysis at a point of ambiguity in an input sentence, and later discovers that subsequent words of the input don't fit into the structure it was building (Fodor and Inoue, 1998). The parser's subsequent efforts at correcting the error are evidenced by prolonged gaze times, increased reaction times, lower rates of recognition, and distinctive neural signatures (Bornkessel-Schlesewsky and Schlesewsky, 2009). Garden path phenomena are arguably best explained by parsing principles such as Minimal Attachment and Late Closure, which dictate that incoming material will be incorporated into the most recent phrase or clause of the existing MPM, and in such a way as to minimize the number of nonterminal nodes (Fodor and Inoue, 1998). These principles are not themselves mentally represented; rather, they are *true of* linguistic processing, in the way that the principles of celestial mechanics are true of our solar system. But importantly, such principles make ineliminable reference to MPMs, and would be explanatorily barren without the reality of that posit. Thus, a realist attitude toward MPMs is warranted.

§3. Sellars and the Explanatory Strategy of Cognitive Science

Talk of mental representations is likely to trigger philosophical suspicions, particularly when the representational contents mark out the unfamiliar categories of a technical field, and when it is unclear that the creatures to whom they are ascribed institute social norms that warrant the attribution of the corresponding concepts. What, then, is it to posit a "mental representation" in psycholinguistics? To answer this, let us focus on the case of MPMs, and correlate the principal tenets of Sellars' treatment of mental representation with some of the core assumptions of psycholinguistic research.

Sellars took causal efficacy to be the mark of reality, and he placed an emphasis on causal *explanation*, both in the context of ontic decision and in his views on the aims of science (deVries, this volume). Correspondingly, psycholinguists hold that MPMs are *causal intermediaries* between the physiological registration of linguistic stimuli, on the one hand, and one's

comprehension of the linguistic input, on the other. Positing these states helps *to explain behavior* (e.g., patterns of eye fixation or reaction times) even in the absence of detailed knowledge of the underlying neural mechanisms.

Mental representation are, for Sellars, functional states, which means that an account of their nature will consist in a specification of their relations to (i) the environment, (ii) behavior, (iii) one another, (iv) underlying mechanisms, and (v) the character of the normative relations that obtain between these things (e.g., teleological or social normativity). Correspondingly, psycholinguists design experiments that crucially rely on an explicit and principled specification of all of these parameters: (i) the stimulus set, (ii) the behavioral variables, (iii) the information-processing model, (iv) the neural implementation model, (v) the learning or evolutionary model. This requires them to think of MPMs as states that reliably covary with the syntactic structure of the stimulus, and whose ability to do so is their *function*, grounded in either natural selection or learning.

Sellars' functional-role semantics (SRLG, ITSA, MFC, MEV) holds that the representational properties of any internal state are exhaustively determined by all three components of its "long-arm" functional role—the environmental conditions under which it is typically elicited, the causal relations it typically bears to other internal states, and the behaviors that it disposes a creature to perform.[5] Contrary to behaviorist, covariationist, and information-theoretic accounts of representation, Sellars held that none of the three factors is individually sufficient; all three are individually necessary and jointly sufficient for something to have representational content (cf. Brandom, 1994: ch. 2; 2009: ch. 7). This contrasts with fully internalist accounts that take cognitive science to be engaged in "individualist" inquiries, which deal solely with the intrinsic properties of the mind/brain, irrespective of the relations that it bears to the environment (Chomsky, 2000).[6] But although some states enjoy more direct causal relations to the environment and to behavior than others, Sellars holds that such causal relations are not by themselves sufficient for an internal state to have representational properties (SRLG). A representation must also enter into computational or inferential relations with other representations. Even a perceptual judgment has its representational properties partly in virtue of its role as a premise in inference and as a guide to action.[7] This stands in opposition to fully externalist accounts, which depict representation as a natural relation between the mind/brain and the environment (Fodor, 1990). On Sellars' account of the pragmatics of representational ascriptions (MFC), to say of some state, S, that it represents X is to classify S as belonging to a particular type (i.e., as playing a distinctive role in the creature's cognitive and behavioral economy). The claim that S represents X does not entail that S bears some particular relation to X—the same unique relation in all cases. There are, of course, many important causal and nomological relations between a creature's psychological states and its environment. And on Sellars' view, these relations constitute two-thirds of an account of those

states' representational properties. Nevertheless, no direct mind-world relation is exhaustive of a state's having the representational properties that it does (cf. Brandom, 1994: 325).

One can imagine being strict about the functional role that an item must play in order to count as a representation of, for example, dogs or noun phrases—what Sellars, employing his famous dot-quotation device, would call a •dog• or a •noun phrase•. But Sellars wisely avoided insisting on strict criteria for the individuation of functional roles, leaving the matter up to the conversational context and the purposes of the attributor. "The criteria which an item must satisfy to be an •or• are a matter of its functioning, *in respects deemed relevant*" (MFC, p. 428, italics added). Such criteria "are flexible and context dependent. What counts as an •or• in one classificatory context may be classified as *like* an •or• in another" (MFC, p. 437, fn. 12).[8] As Seibt (2009: 243, fn. 4) points out, "Sellars expressly introduces an 'extended interpretation of the dot-quoting device' to allow for inclusion of non-linguistic functional analogues (MEV §§76–7: 340)." Mental phrase markers are ripe for inclusion in this extended class. To characterize an MPM as being a •This is a noun phrase• is to note the similarity of its function within the parsing mechanism to the functional role of the sentence 'This is a noun phrase' in ordinary English. True, the similarity is very partial, excluding a wide variety of inferences that are possible only at the personal level (§§3–4). But it is nevertheless informative, for the parsing mechanism can be counted on, in a wide range of circumstances, to draw the same "inferences" from this MPM as a competent linguist would draw from the corresponding sentence. This is nowhere clearer than in computational models in the "parsing as deduction" tradition, which treat parsing as a quasi-inferential process of deriving theorems from premises about the linguistic input and axioms that constitute the grammar (Johnson, 1989).[9] In such models, the relations between MPMs resemble rational inferences—reasonable steps toward a cognitive goal—corresponding to Sellars' language-internal moves (SRLG).

Talk of rationality and goals raises a further point. Sellars' account of representation stresses the normativity inherent in ascriptions of representational states. Representation is a kind of *purport* (Brandom, 1994), which is always in principle subject to failure—what we call *error* or *mis*representation. Correspondingly, psycholinguists think of MPMs as states that can miscategorize the input—as in the case of garden-path sentences and garden-variety misinterpretations. Indeed, patterns of such errors constitute the principal explananda for psycholinguistic theorizing. Detecting signs of such errors, either in behavior or in the brain, is a basic concern for experimental design (Bornkessel-Schlesewsky and Schlesewsky, 2009).

Finally, consider the interplay between predication and singular reference (MEV). Predication implements semantic generality—a one-to-many relation between representations and the particulars they represent. Reference allows for semantic singularity—treating one and the same particular

as belonging to two different categories by applying two predicates to instances of the same singular term. Both of these semantic functions can be discerned in the psycholinguist's assumptions about MPMs. All psycholinguistic models of the language processing system build in a kind of semantic generality—the ability to treat acoustically diverse stimuli as belonging to the same category (e.g., representing both 'my book' and 'a spritely ostrich' as noun phrases). Less obviously, but equally importantly, the models can treat one and the same stimulus as belonging to two different classes. For instance, in processing sentence (8), the parser first categorizes 'the kids' as a noun phrase, and only later learns that it is also the syntactic object of 'observed' (a relation made explicit by the index notation).[10]

(8) These are [the kids]$_i$ who were only very rarely observed ___$_i$ in the classroom.

It is no accident that there are numerous and substantial correspondences between how Sellars thinks of mental representations and how psycholinguists characterize them in their workaday practices. One of Sellars' lasting contributions to philosophy of mind was his explicit articulation of what we now think of as functionalism. For Sellars, this was the view that psychological states are good theoretical posits, whose nature is exhausted neither by what is revealed in introspection, nor by behavioral dispositions and neural markers, but rather, by their functional relations to stimuli, behavior, and one another. Historically, it was Sellars' opposition to mind-body dualism and reductive behaviorism that helped to set the stage for the next generation of thinkers—Chomsky (1965), Fodor (1968; 1983), Dennett (1978), and others—to usher in the paradigm shift that we now think of as the cognitive revolution.[11] These theorists inherited the fruits of Sellars' labor, including his demolition of positivist and instrumentalist assumptions in the philosophy of science, his scientific realism, his naturalism, his functionalism, and his insistence on the importance of explanation as against mere description. This latter point is echoed distinctly in Fodor's principal objections to reductive and analytic forms of behaviorism (Fodor, 1968; Fodor, Bever, and Garrett 1974), which he argues to be incapable of explaining cognitive processes. It is equally evident in Chomsky's call to treat syntactic theorizing as an explanatory enterprise, rather than a merely taxonomic exercise governed by so-called "discovery procedures" (Chomsky, 1965). It is fair to say, then, that Sellars was one of the principal architects of the conceptual foundations on which cognitive science was built.

§4. Subpersonal States

Contemporary cognitive science, of which psycholinguistics is a particularly well-developed branch, presumes a multilevel structure of description and explanation. High-level descriptions of persons and their gross

psychological capacities (vision, planning, memory, etc.), are subserved by lower-level descriptions of one or another type of computational operation, which are in turn subserved by even lower-level descriptions couched in the language of neurobiology, molecular biology, and ultimately, particle physics (Dennett, 1978). Whereas folk psychology describes and explains behavior in terms of states that are ascribed to a person, cognitive science offers *sub*personal explanations, which ascribe states to the information-processing mechanisms that jointly constitute a person—what are sometimes called "mental organs" or "modules" (Fodor, 1983).

This encourages a vision in which a community like ours has internalized the folk-psychological lessons of the mythical genius Jones and has put itself in a position to refine their understanding of themselves by incorporating the insights of a new genius—the not-so-mythical Noam. But although Sellars would no doubt have been friendly to the idea of adding levels of description to our theorizing, he would have insisted that we be clear on what commitments we take on board in offering accounts that appeal to subpersonal states and operations. For, although there are compelling grounds for thinking of some subpersonal states as *representations* (§2), it would be a mistake to uncritically assimilate them to personal-level beliefs, desires, and intentions. The Myth of Noam, though modeled on the Myth of Jones, must come with a commentary. In this section, I draw on Sellars' conceptual toolkit to articulate the main differences between states like the psycholinguist's MPMs and the denizens of our Manifest Image folk psychology.

Consider first an issue pertaining to normativity. Sellars held that meaning, representation, intentionality, and related phenomena are "fraught with 'ought'" (TC/SPR: 212). But a different flavor of normativity is involved in the ascription of folk-psychological states and processes (e.g., intentions and planning) than in the in the ascription of computational states and operations—data structures and algorithms. A long-time proponent of Sellars' work, Daniel Dennett, has captured this idea neatly in his distinction between three "stances" that one can adopt in describing, explaining, and interacting with natural systems (Dennett, 1987).

Adopting the physical stance, one takes as one's target a physical object, which is presumed to have mass, velocity, temperature, and the like. Presumably, no normative assumptions are presupposed in the ascription of such attributes. One level up is Dennett's design stance, which characterizes its targets as well-functioning designed systems. Taking this stance toward a creature involves thinking of it as an aggregate of interacting purposeful mechanisms, each of which has the function of performing a specialized task. It's from the design stance that we attribute subpersonal states to the human sentence processing mechanism—a gadget whose "job" is to classify incoming linguistic input. The normative assumptions that are in play here are broadly teleological, grounded in either learning or natural selection. Finally, one can take on even stronger normative commitments and ascend to the intentional stance, treating one's target as a rational agent,

who perceives and remembers what she ought to, in order to make reasonable inferences that lead to intelligent action.

The difference between the teleological normativity of the subpersonal design level and the discursive normativity involved in a personal-level intentional ascription is reflected in the kinds of criticisms we make in different circumstances. Consider the following example: Inspecting a used car that you are considering purchasing, you discover that the salesman has failed to mention an important defect. "Did he forget? Or did he think it was irrelevant? He probably concealed it intentionally, in which case he's not stupid, but outright malicious. But even then, did he not expect me to notice?" Such suspicions of moral or epistemic failure do not arise, by contrast, when we regard cases of visual illusions or their linguistic analogues, garden path sentences. Though such cases plainly involve error and misrepresentation, the source of the error is not a failure of rationality or moral virtue, but a malfunction in a subpersonal mechanism, whose heuristic operating assumptions lead it astray when presented with a circumscribed class of inputs. No one can be called "stupid" for falling prey to the Müller-Lyer illusion, or for failing to comprehend sentences (2)-(7) on a first pass.

The normative contrast can even be seen in a single case, as when a friendly shopkeeper gives incorrect change for one's morning coffee.[12] If one grasps at an intentional explanation, one must posit either malicious motives or false beliefs. Having ruled out the former on grounds of trust, one searches in vain for a suitable set of false beliefs to attribute. Given the rationality assumption, it's simply not plausible that the shopkeeper believes that the coffee costs more today than it did yesterday, or that subtracting 50¢ from $1 yields 20¢, or that the two dimes he gave you are actually quarters. One intuitively descends to the design level, positing some dimly imagined subpersonal mechanism—perhaps a bleeding of information between the registers in his mental math module, or an error in his early visual processing concerning the size of the coins. One can descend to an even lower rung, positing noisy or leaky neuronal channels. Teleological normativity gradually gives out when the downward descent hits the levels of molecules, atoms, and quarks.

A different way of casting what it means to "adopt a stance" is in terms of ontological commitment. Sellars took theoretical postulation of *individuals* to be a characteristic achievement of the Scientific Image. But not all theoretical posits are individuals; some are *states* of already-familiar individuals. As O'Shea (2012: p. 195, fn. 22) points out, Sellars held that Jones' mythical postulation of folk-psychological states did not cross the line into the Scientific Image.

> Perhaps the most important point is that what [Jones'] theory postulates in the way of new entities are processes and acts rather than [new] individuals. In this sense, it remains within the manifest image. Persons remain the basic individuals of the system.
>
> (SK II, §55)

Cognitive science goes beyond Jonesian folk psychology in postulating mental organs as its "basic individuals"—modular elements, on a par with the hearts, cells, and chromosomes of biology (Chomsky, 2000). Ontological commitment to such entities brings us squarely into the Scientific Image.

Another mark of subpersonal states is their inexpressibility in speech. Mythical Jones took as his starting point the overt thinkings-out-loud of his Rylean fellows, characterized in semantic and pragmatic terms (EPM). He then posited short-term propensities to produce them; personal-level states then show up as the causal grounds of those propensities, which, as a surplus, serve to causally explain rational *non*verbal behavior. Mythical Noam, by contrast, infers subpersonal MPMs from fine-grained behavioral data—eye-fixation times, recall tasks, and the like—as well as neurophysiological markers. Although MPMs do play a causal role in language production, they do not meet other conditions that constitute Sellars' overall account of verbal expression (LTC). Their representational contents, which pertain to grammatical relations, do not coincide with the meanings of the speech acts whose production they facilitate. Nor does the illocutionary force of those speech acts in any way reflect the mental attitudes of the subpersonal states that facilitate them.

A related feature of subpersonal states is their permanent inaccessibility to consciousness. Mythical Jones trained his fellows to report one another's thoughts, and eventually to ascribe them to oneself, on the basis of either overt behavior—notably, one's own speech—or one's own verbal dispositions (EPM). Drawing on Sellars, Rosenthal (2005) argues that when this kind of self-ascription becomes automatic and second-nature, one comes to have the kinds of higher-order thoughts necessary for both consciousness and introspection. But such a process cannot have occurred in the case of subpersonal states. Given the fine-grained nature of the data that Mythical Noam appeals to in positing MPMs, and given their inexpressibility in speech, such states would not have become routinely reportable, and hence could not come to be ascribed to oneself in an automatic fashion.

Rosenthal (2005) takes Sellars' discussion in EPM as an inspiration for his higher-order thought (HOT) theory of consciousness. If that theory is on the right track, then it would explain the oft-noted fact that subpersonal states are invariably nonconscious and unavailable for introspective report. Indeed, no other explanation seems credible. Consider the usual candidates: Psychological defense mechanisms, such as repression, keep troublesome thoughts and desires out of consciousness, but would have no cause to suppress emotionally neutral states like MPMs. Nor are MPMs dormant, like long-term memories; they're occurrent states, accessed in real-time information processing. Last, although brief or degraded presentation of stimuli can render a perceptual judgment nonconscious, this is implausible as an explanation of why MPMs never occur consciously, for they are activated by ordinary linguistic inputs, and participate in a processing stream that

ultimately gives rise to a conscious experience of comprehension. Sellars' views thus contain not only the insight that personal-level states can occur nonconsciously—indeed, that this is their default state—but also the germ of a satisfying account of how consciousness comes about. Better still, it can explain what other accounts merely stipulate—viz., that subpersonal states are invariably nonconscious. It thus serves as a useful corrective to the frequent conflation of the personal/subpersonal distinction with the conscious/nonconscious distinction.

Sellars held that personal-level states come to be what they are by getting caught up in a vast inferential network. Their intentional contents are properly thought of as conceptual, on account of the inferential norms that are constitutive of discursive practice. By contrast, subpersonal states are not "inferentially integrated," in the sense that we cannot draw inferences whose premises or conclusions are MPMs (Stich, 1978). Such states are thus not "in the space of reasons." This fits well with the fact that they are inexpressible in speech, for that would explain why they can never be played as counters in the game of giving and asking for reasons. Still, just as it would be a mistake to hastily assimilate MPMs to personal-level states, it would be equally wrong to see them as "mere causes" with no psychological properties. For, although their contact with personal-level states is limited to facilitating linguistic comprehension, they are nevertheless enmeshed in their own *subpersonal* system of quasi-inferential entry, exit, and internal moves, which has been successfully modeled by computational linguists (Johnson, 1989; Pereplyotchik, 2017). The limited number of pieces and moves in this game is precisely what makes it computationally tractable (in principle). By contrast, Brandom (2008) has argued that there are potentially insurmountable problems facing any algorithmic model of nonmonotonic material inference.[13]

Summing up, subpersonal states have only *some* of the features characteristic of personal-level beliefs and desires. They bear systematic relations to the environment, to behavior, and to one another. This makes it both reasonable and useful to think of them as representations whose interactions resemble inferences between propositional attitudes. Doing so allows us to abstract away from largely unknown neural mechanisms and to rationalize subpersonal mechanisms, seeing them as engaged in purposeful activities that involve reasonable steps toward a cognitive goal. But while it's tempting to assimilate such states to propositional attitudes (Fodor, 1983; 1987), this would be a mistake. Sellars' framework for thinking about mental phenomena can be marshaled both in distinguishing subpersonal states from their personal-level counterparts, and in articulating the conceptual connections between their distinguishing features—their inexpressibility in speech, absence from consciousness, inferential isolation, susceptibility to computational modeling, and teleological normative status.

§5. Syntactic Rules as Auxiliary Positions

In their book, *Knowledge, Mind, and the Given*, deVries and Triplett point out that "[t]he definitive treatment of the relation between Chomsky and Sellars has yet to be written" (p. 61, fn. 3). A comprehensive treatment would have to make mention of the fact that Chomsky published reviews of both SRLG and ITSA early in his career (Chomsky, 1957a,b). Neither scholar had yet achieved the status of a major figure in the field, which perhaps goes some way toward explaining why these reviews appear rushed, displaying little care for the subtleties of Sellars' position, and even less patience for his major conclusions. This is unfortunate, in part because SRLG contains a distinction that I believe would have clarified subsequent debates concerning Chomsky's views on the "psychological reality" of grammar. Sellars distinguishes between two types of position in a language game—ordinary and auxiliary. Chomsky notes the distinction, but does not make anything of it. I shall argue that the distinction between what I've been calling mental phrase markers (MPMs) and mental syntactic rules (MSRs), which permeates Chomsky's views on language, is a subpersonal analogue of Sellars' distinction between ordinary and auxiliary positions. This, in turn, bears on how we view the debate concerning the psychological reality of MSRs.

In the following passage from SRLG, Sellars introduces the ordinary/auxiliary distinction:

> [In addition to observation and inference there is] a third way of properly coming to be at a position. Here one comes to be at certain positions without having moved to them from other positions (in which position it resembles observation), and without having made a language entry transition (in which respect it resembles inference). The positions in question are 'free' positions which can properly be occupied at any time if there is any point to doing so. Obviously what I have in mind are the sentences the status of which, when used in a rule obeying manner, is specified as that of 'primitive sentence' (i.e. as unconditionally assertable) by a rule of the metalanguage. (Thus, 'All A is B' might be specified as a primitive sentence of language game L.) Are such sentences properly called positions? Their 'free' status and their 'catalytic' function make them a class apart, yet it is less misleading to call them positions than it would be to call sensations positions. Let us call them 'auxiliary positions'.
>
> (*SPR*, p. 330)

The characteristic features of auxiliary positions include their *propriety*, their *generality*, and their "catalytic function." Sellars goes on to elaborate on the latter notion, pointing out another crucial feature of auxiliary positions—their *dispensability*.

> [A] language game which contains the auxiliary position 'All A is B' makes possible the syllogistic from 'This is A and All A is B' to 'It is B'. An alternative way of going from 'This is A' to 'It is B' would exist if the game included a direct move from positions of the form '. . . is A' to positions of the form '. . . is B'. We thus notice a certain equivalence between auxiliary positions and moves. We also notice that while it is conceivable that a language game might dispense with auxiliary positions altogether, though at the expense of multiplying moves, it is not conceivable that moves be completely dispensed with in favour of auxiliary positions. A game without moves is Hamlet without the Prince of Denmark indeed!
>
> (*SPR*, p. 330; cf. Fodor, 1987: pp. 21–26)

Sellars' remarks in this passage are developed more fully in MEV, where he contrasts Humean and Aristotelian representational systems (RSs). A Humean RS is one that lacks logical vocabulary, including quantifiers and connectives. The internal transitions in a Humean RS are all like material inferences—good or bad in virtue of their nonlogical components. By contrast, an Aristotelian RS is capable of representing lawlike relations *between* kinds, using universal quantifiers and conditionals (or some equivalent apparatus), thus allowing its inferences to be good or bad in virtue of their logical form. Though he did not put it this way in SRLG, an Aristotelian RS makes use of auxiliary positions, whereas a Humean RS is one that dispenses with them.

We can now apply Sellars' ordinary/auxiliary distinction to the subpersonal case of language processing. MPMs are subpersonal analogues of personal-level thoughts—particularly perceptual judgments—in that they are occurrent states whose function is to represent the syntactic properties of incoming linguistic stimuli. To token an MPM is to briefly occupy a subpersonal *ordinary* position. By contrast, MSRs are best seen as subpersonal analogues of *auxiliary* positions—i.e., standing states of the human parsing mechanism that embody the rules or principles of a person's language (Pereplyotchik, 2017). To illustrate the use and plausibility of this analogy, let's run through two parallel types of psychological explanation.

At the personal level, we tell stories such as the following: "Ever since he was a child, Jeff has believed that spiders are dangerous. So when he just now saw several spiders crawl out of a box in his attic, he reasoned that there are probably other dangerous bugs in the box." In this story, a perceptual judgment regarding the presence of spiders (an ordinary position) interacts with a standing belief (an auxiliary position) to yield an occurrent thought (another ordinary position) as a conclusion. The process of syntactic parsing, though it takes place subpersonally, has roughly the same structure. Consider, for instance, what happens when the human language processor encounters an ambiguous sentence-opening.

(9) Have the soldiers . . . [*possible continuations*: . . . eaten? . . . fed!]

The parser houses two standing belief-like states—two MSRs—that embody the following two principles: (i) a sentence-initial auxiliary verb (e.g., 'Have') serves to introduce a question; (ii) a subsequent noun phrase (e.g., 'the soldiers') will be the subject of the clause. These are auxiliary positions in the subpersonal game of sentence parsing. As a result of registering the presence of the word 'Have' (an ordinary position), and occupying the abovementioned auxiliary positions, the parser constructs an MPM (another ordinary position) that represents 'Have' as an inverted auxiliary verb. From this MPM and the two standing MSRs, the parser "concludes"—sometimes incorrectly—that the overall sentence will be a question, and hence that the subsequent noun phrase will be the subject of the clause.

MSRs are *general* in that they are applicable in indefinitely many cases. For instance, a structural MSR such as S → NP VP is a recipe for (de)composing a wide range of sentences. MSRs are also arguably *proper*, though in a teleological rather than a social sense (§3).[14] The question that we must now face is whether MSRs have the remaining feature characteristic of auxiliary positions—*dispensability*. This question is at the center of the debate concerning the "psychological reality" of grammar. We can clarify this debate by casting the central question in terms of Sellars' conceptual framework: Is the human parsing mechanism a (subpersonal) Humean RS or an Aristotelian one? Do MSRs serve a "catalytic function," causally mediating between various MPMs? Note that the psychological reality of MPMs—subpersonal ordinary positions—is not in dispute. While it is conceivable that a parsing mechanism might dispense with MSRs altogether, a parser without computational operations on MPMs is Hamlet without the prince of Denmark indeed!

Having clarified the question by casting it in Sellarsian terms, I will not pursue it further here. I have argued elsewhere (Pereplyotchik, 2017) that we have, at present, no evidence for supposing that the human parsing mechanism is an Aristotelian system. As far as we know today, the hypothesized mental grammar is nothing more than a kind of embodied *procedural* knowledge—a set of systematic dispositions to move directly from some types of MPM to others. I close this discussion by drawing on another distinction from Sellars in sketching the residual explanatory value of MSRs.

Though philosophers sometimes apply the term 'belief' to any assertoric propositional attitude, Sellars pointed out (LTC) that this usage elides an important distinction between *occurrent* assertoric attitudes and *dispositions to have* occurrent assertoric attitudes. The dispositional/occurrent distinction is not only significant for the purposes of theory construction, but is also clearly marked in folk-psychological explanation.[15] On the more careful usage that Sellars recommends, thoughts are states that we come to be in *occurently*, as a result of perception or inference, whereas beliefs are *dispositions* to have certain occurrent thoughts when the occasion arises, or to process various occurrent thoughts in specific ways.[16]

Lacking auxiliary positions, a Humean RS can be said to have general beliefs only in the sense that it has procedural dispositions to move from one

thought to another. For instance, its belief that all red strawberries are ripe is nothing more than its disposition to infer 'This is ripe' from 'This is a red strawberry', 'This is red' from 'This is a ripe strawberry', and so on. Although it would be a mistake to take such beliefs as causally active episodes, it is nevertheless illuminating to describe the RS as *having* beliefs. The procedural nature of these beliefs in no way hampers our ability to explain and predict behavior of the RS by classifying it in intentional or semantic terms. And lacking any practical way of characterizing the RS in neurophysiological terms, the predictive and explanatory leverage that we derive is enormous—arguably indispensable (Dennett, 1987). Given the analogy between the belief/thought contrast at the personal level and the MSR/MPM contrast at the subpersonal level, we can likewise see the value of talking in terms of MSRs. Even if they are nothing more than dispositions to token occurrent MPMs, this would not, I contend, impugn their explanatory value for psycholinguistics.

Acknowledgments

My sincere thanks to Jacob Berger, Bill deVries, Kevin Fink, and Carl Sachs for their helpful comments on previous drafts. Earlier versions of this paper were presented at a meeting of the Wilfrid Sellars Society at the American Philosophical Association, Eastern Division (2014), and at the Sellars in a New Generation conference at Kent State University (2015). I am grateful to the audiences for their questions and comments.

Notes

1 One might dispute this claim, citing passages like this one from SM: "But *no* picture of the world contains *as such* mentalistic expressions functioning *as such*. The indispensability and logical irreducibility of mentalistic discourse is compatible with the idea that in *this* sense *there are no mental acts*" (SM, V, 78: 143). (Thanks to Bill deVries for pressing this objection.) How one responds will depend on what one takes to be Sellars' view of the relation between "picturing" and truth, or between the Manifest and Scientific images. For discussion of these issues, see the papers in section III of this volume.

2 Syntacticians in the early days of generative grammar talked of syntactic *rules*, but in the 1990s began to favor talk of syntactic *principles* (Chomsky, 1995). This terminological shift is important (Pereplyotchik, 2017), but it raises issues that are beyond the scope of the present discussion, so I will continue to use the outdated but familiar term, 'rules'.

3 For a review of the innateness controversy, see Pinker (1994) and Cowie (2010). Sellars' views leave room for innateness. In MEV, he writes: "a primitive [representational system] is also an innate endowment of human beings. The concept of innate abilities to be aware of something as something, and hence of prelinguistic awarenesses is perfectly intelligible" (§57).

4 Contemporary syntactic theory has moved well beyond this type of grammar, and the structural analyses that more sophisticated grammars generate are demonstrably superior to (1) in many respects. I set aside these differences for present purposes; see Pereplyotchik (2017) for details.

5 Versions of functional-role semantics have been advanced by Dennett (1969, 1978, 1987), Harman (1975, 1987), Field (1978), Churchland (1979), Brandom (1994, 2008, 2009), and others.
6 Chomsky's individualism/internalism about representation bears a close resemblance to the view that Brandom (1994) labels 'hyperinferentialism'.
7 Following Sellars, Dennett (1969: 76–8) argues that neural event's being reliably elicited by food is not sufficient for its being a representation of food. In order to be a representation of food, a state must also cause the right sorts of representations downstream, culminating behaviors that are, in some sense, appropriate specifically to food (for that creature) (cf. Dennett 1978, 1987). Similarly, Churchland (1979: ch. 2) presents cases where focusing on the state's causal origin would result in gross misinterpretation, but focusing on inferential role yields the right result. Brandom (2009: ch. 7) makes this point in his discussion of the important difference between reliable *detection* and genuine *representation*.
8 Quine (1960, §45) makes a similar point: "An indirect quotation we can usually expect to rate only as better or worse, more or less faithful, and we cannot even hope for a strict standard of more and less; what is involved is evaluation, relative to special purposes, of an essentially dramatic act." See Hicks (this volume) for discussion of the differences between Sellars and Quine on this issue.
9 The axioms can be traded in for inference rules (Pereplyotchik, 2017). I discuss Sellars' conception of this trade-off in §5.
10 The 'recognize *as*' locution is indispensable for psychological explanation. For instance, Fodor (1989) points out that "antecedents typically cannot, when they are encountered, be recognized *as* antecedents" (p. 197, emphasis on 'as' in the original). She then explains a particular type of garden-path effect by noting that recognition of an antecedent *as such* takes additional time, beyond the time required for recognizing the phrase to which the parser in effect assigns that predicate.
11 "It rained for weeks and we were all *so* tired of ontology, but there didn't seem to be much else to do. Some of the children started to sulk and pull the cat's tail. It was going to be an *awful* afternoon until Uncle Wilifred [*sic*] thought of Mental Representations (which was a game that we hadn't played for *years*) and everybody got very excited and we jumped up and down and waved our hands and all talked at once and had a perfectly *lovely* romp" (Fodor, 1985: 76).
12 This example is adapted from Dennett (1987).
13 For similar arguments, see Fodor (1983) and Putnam (1988).
14 One might object that there are social dimension to the norms governing the grammar that one employs. Although I do not believe that the normativity involved in the ascription of subpersonal states is social or discursive, I would grant that there is a complex interplay between the social and teleological flavors of normativity. Getting clear on the relations between them would go a long way toward developing a satisfying account of both the evolution of language in a species and the acquisition of language in an individual.
15 Q: "Why did the student get this question wrong on the math exam?" A: "He must have thought that the square root of 4 can only be a positive integer." Q: "But if you had asked him straight out, he would have told you that it can be negative too." A: "Yes, I'm sure he *believes* that it can, but he must have not *thought* of it at the time."
16 Similar remarks can be made with respect to non-assertoric states. There is, for instance, a distinction between standing preferences and occurrent *wants* or *urges*. Q: "Do you prefer chocolate ice-cream or vanilla?" A: "I prefer chocolate, but I don't want any right now; I just ate two portions of cake and I have no urge to eat any more dessert."

References and Further Reading

Bornkessel-Schlesewsky, I. and Schlesewsky, M. (2009). *Processing Syntax and Morphology: A Neurocognitive Perspective*. Oxford: Oxford University Press.

Brandom, R. (1994). *Making It Explicit*. Cambridge, MA: Harvard University Press.

Brandom, R. (2008). *Between Saying and Doing*. New York: Oxford University Press.

Brandom, R. (2009). *Reason in Philosophy*. Cambridge, MA: Harvard University Press.

Chomsky, N. (1957a). "Review of 'Is there a Synthetic a Priori,' by Wilfrid Sellars," *The Journal of Symbolic Logic*, 22 (4), pp. 402–403.

Chomsky, N. (1957b). "Review of 'Some Reflections on Language Games,' by Wilfrid Sellars," *The Journal of Symbolic Logic*, 22 (4), p. 403.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge, MA: Cambridge University Press.

Churchland, P. (1979). *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.

Cowie, F. (2010). "Innateness and Language," in *Stanford Encyclopedia of Philosophy*, E. Zalta (ed.). http://plato.stanford.edu/archives/sum2010/entries/innateness-language/

Dennett, D.C. (1969). *Content and Consciousness*. London: Routledge & K. Paul.

Dennett, D.C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, MA: MIT Press.

Dennett, D.C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett, D.C. (2010). "The Evolution of 'Why'," in *Reading Brandom: On Making It Explicit*, B. Weiss and J. Wanderer (eds.). New York: Routledge, pp. 48–62.

DeVries, W., ed. (2009). *Empiricism, Perception, Knowledge, Normativity, and Reason: Essays on Wilfrid Sellars*. Oxford: Oxford University Press.

DeVries, W. and Triplett, T. (2000). *Knowledge, Mind, and the Given: A Reading of Sellars' "Empiricism and the Philosophy of Mind"*. Cambridge, MA: Hackett Publishing. [KMG].

Field, H. (1978). "Mental Representation," *Erkenntnis*, 13, pp. 9–61.

Fodor, J.A. (1968). *Psychological Explanation: An Introduction to the Philosophy of Psychology*. New York: Random House.

Fodor, J.A. (1983). *The Modularity of Mind*. Cambridge: MIT Press.

Fodor, J.A. (1985). "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum," *Mind*, New Series, 94 (373), pp. 76–100.

Fodor, J.A. (1987). *Psychosemantics*. Cambridge: MIT Press.

Fodor, J.A. (1990). *A Theory of Content, and Other Essays*. Cambridge: MIT Press.

Fodor, J.A., Bever, T., and Garrett, M. (1974). *The Psychology of Language*. New York: McGraw Hill.

Fodor, J.D. (1989). "Empty Categories in Sentence Processing," *Language and Cognitive Processes*, 4, pp. 155–209.

Fodor, J.D. (1995). "Comprehending Sentence Structure," in *An Invitation to Cognitive Science: Volume 1, Language*, L. Gleitman and M. Liberman (eds.). Cambridge, MA: MIT Press, pp. 209–246.

Fodor, J.D. and Inoue, A. (1998). "Attach Anyway," in *Reanalysis and Sentence Processing*, J.D. Fodor and F. Ferreira (eds.). New York: Kluwer Academic Publishers, pp. 101–141.

Harman, G. (1975). "Language, Thought, and Communication," in *Language, Mind, and Knowledge*, K. Gunderson (ed.). Minneapolis: University of Minnesota Press, pp. 270–298.

Harman, G. (1987). "(Nonsolipsistic) Conceptual Role Semantics," in *New Directions in Semantics*, Ernest LePore (ed.). London: Academic Press, pp. 55–81.
Johnson, M. (1989). "Parsing as Deduction: The Use of Knowledge of Language," *Journal of Psycholinguistic Research*, 18 (1), pp. 105–128.
Jurafsky, D. and Martin, J.H. (2008). *Speech and Language Processing: An Introduction to Speech Recognition, Computational Linguistics, and Natural Language Processing (2nd edition)*. Englewood Cliffs, NJ: Prentice Hall.
O'Shea, J. R. (2012). "The 'Theory Theory' of Mind and the Aims of Sellars' Original Myth of Jones," *Phenomenology and the Cognitive Sciences*, 11 (2), pp. 175–204.
Pereplyotchik, D. (2015). "Review of *Empiricism, Perception, Knowledge, Normativity, and Realism: Essays on Wilfrid Sellars*," *Journal for the History of Analytic Philosophy*, 3 (8), pp. 1–19. Retrieved from https://jhaponline.org/jhap/article/view/2728/2495
Pereplyotchik, D. (2017). *Psychosyntax: The Nature of Grammar and Its Place in the Mind*. New York: Springer.
Phillips, C. and Lewis, S. (2013). "Derivational Order in Syntax: Evidence and Architectural Consequences," *Studies in Linguistics*, 6, pp. 11–47.
Pickering, M.J. and Ferreira, V.S. (2008). "Structural Priming: A Critical Review," *Psychological Bulletin*, 134 (3), pp. 427–459.
Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: William Morrow and Co.
Putnam, H. (1988). *Mental Representation and Reality*. Cambridge, MA: MIT Press.
Quine, W.V. (1960). *Word and Object*. Cambridge: MIT Press.
Rosenthal, D.M. (2005). *Consciousness and Mind*. Oxford: Oxford University Press.
Scharp, K. and Brandom, R.B., eds. (2007). *In the Space of Reasons*. Cambridge, MA: Harvard University Press. [ISR].
Seibt, J. (2009). "Functions between Reasons and Causes: On Picturing," in *Empiricism, Perception, Knowledge, Normativity, and Realism*, W. deVries (ed.). New York: Oxford University Press, pp. 247–281.
Sellars, W.S. (1953a). "Is there a Synthetic *A Priori*?" first published in *Philosophy of Science*; reprinted with revisions in *American Philosophers at Work: The Philosophic Scene in the United States*, Sidney Hook (ed.). New York: Criterion Books, 1956, pp. 135–159. Reprinted in SPR.
Sellars, W.S. (1953b). "Inference and Meaning," *Mind*, 62 (247), pp. 313–338. Reprinted in ISR.
Sellars, W.S. (1954). "Some Reflections on Language Games," *Philosophy of Science*, 21, pp. 204–228. Reprinted with extensive revisions in SPR and ISR.
Sellars, W.S. (1956). "Empiricism and the Philosophy of Mind," in *Minnesota Studies in the Philosophy of Science*, vol. I, H. Feigl and M. Scriven (eds.). Minneapolis: University of Minnesota Press, pp. 253–329. Reprinted in SPR and KMG.
Sellars, W.S. (1962a). "Truth and 'Correspondence'," *The Journal of Philosophy*, 59 (2), pp. 29–56. Reprinted in SPR.
Sellars, W.S. (1962b). "Philosophy and the Scientific Image of Man," in *Frontiers of Science and Philosophy*, R. Colodny (ed.). Pittsburgh, PA: University of Pittsburgh Press, pp. 35–78. Reprinted in SPR and ISR.
Sellars, W.S. (1963). *Science, Perception, and Reality*. London: Routledge & Kegan Paul Ltd, and New York: The Humanities Press, 1963; reissued in 1991 by Ridgeview Publishing Co., Atascadero, CA. [SPR].
Sellars, W.S. (1968). *Science and Metaphysics: Variations on Kantian Themes*. London: Routledge & Kegan Paul Ltd, and New York: The Humanities Press; 1968. Reissued in 1992 by Ridgeview Publishing Co., Atascadero, CA.
Sellars, W.S. (1969). "Language as Thought and as Communication," *Philosophy and Phenomenological Research*, 29, pp. 506–527. Reprinted in ISR.

Sellars, W.S. (1974). "Meaning as Functional Classification," *Synthese*, 27, pp. 417–437. Reprinted in ISR.

Sellars, W.S. (1975). "The Structure of Knowledge: (1) Perception; (2) Minds; (3) Epistemic Principles," in *Action, Knowledge and Reality: Studies in Honor of Wilfrid Sellars*, H.-N. Castaneda (ed.). Indianapolis: Bobbs-Merrill, pp. 295–347.

Sellars, W.S. (1981). "Mental Events," *Philosophical Studies*, 39, pp. 325–345. Reprinted in ISR.

Stich, Stephen P. (1978). "Beliefs and Subdoxastic States," *Philosophy of Science*, 45 (4), pp. 499–518.

Sturt, P., Pickering, M.J., Scheepers, C. and Crocker, M.W. (2001). "The Preservation of Structure in Language Comprehension: Is Reanalysis the Last Resort?" *Journal of Memory and Language*, 45, pp. 283–307.

6 Sentience and Sapience

The Place of Enactive Cognitive Science in Sellarsian Philosophy of Mind

Carl B. Sachs

0. Introduction

It has become something of an orthodoxy amongst philosophers working in the wake of Sellars—especially Brandom and McDowell—that there is a fundamentally important distinction between sapience and sentience. Sapience concerns our ability to make assertions and other normatively governed speech acts with inferentially articulated propositional contents, and to entertain thoughts with analogous structure. By contrast, sentience concerns our ability to discriminate perceptually between motivationally salient stimuli and act in correspondingly appropriate ways.[1] At least in Sellars himself, both sentience and sapience are 'transcendental', in the following highly restricted sense. Transcendental reflection is an inventory of the most general kinds of cognitive capacities and incapacities necessary for the kind of cognitively significant experiences for beings such as ourselves or any being that we are capable of recognizing as being like ourselves.[2] (Those cognitive experiences include but are not limited to knowledge.) As a result of this reflective process, we are led to introduce the distinction between sentience and sapience. However, Sellars should also be read as having as a methodological principle that "transcendental structures must be realized in causal structures" (deVries 2011, pp. 61–62). That is, whatever structures and processes that we posit in transcendental inquiry must be causally implemented by structures and processes that are empirically confirmed and, to the extent possible, consistent with a scientific view of the world, however broadly conceived.

Within this generally Sellarsian framework, I aim to criticize how Sellars conceptualizes the nature and role of sentience in perceptual experience. Sellars maintains that perceptual experience involves both conceptual and nonconceptual content. In contrast to both direct realism and sense-datum theories, Sellars defended a version of *critical realism*—a view that depends heavily on Sellars' interpretation of what Kant called "the productive imagination" (§1). Sensations are not epistemic intermediaries between us and things (*contra* sense-datum theories), but causal intermediaries (*contra* direct realism, which denies that philosophical reflection on experience requires any intermediaries between us and things[3]). Nevertheless, recent

work in embodied cognitive science by Alva Noë, Mark Rowlands, and Anthony Chemero, motivated by the ecological psychology of J. J. Gibson, makes direct realism a more attractive view than critical realists recognize (§2). I will show how Sellarsian critical realism and Gibsonian direct realism can be integrated in what I call *embodied critical realism* (§3).

1. The Role of Embodiment in Sellars' Theory of Perception

Most discussions of Sellars' theory of perception have focused on the role of concepts and sense-impressions in the overall account—thereby generating much discussion over how to read Sellars in light of the debate between conceptualism and nonconceptualism (see O'Shea 2010). However, less attention has been given to the details of how Sellars understands how concepts and sense-impressions are supposed to fit together (with the notable exceptions of Coates 2007 and Levine 2016). This part of the theory is most fully developed in Sellars' late essay, "The Role of Imagination in Kant's Theory of Experience" (IKTE), to which I now turn by taking up the following questions. What role does "the productive imagination" play in Sellars' theory of perceptual experience? What problems is it invoked to solve? What is the role of the perceiver's own body in this theory? And how does the perceiver's body relate to the constraining role of 'sheer receptivity' that sense-impressions play in perceptual takings?

As Sellars sets up the exposition of visual perception, he draws our attention to the fact that our visual experience appears to immediately represent three-dimensional entities in space and time, causally interacting with other such objects, even though what we immediately see *of* those objects is their facing sides. How, then, do we experience the interiors and non-facing sides of objects that do not immediately confront us in visual perception? One answer that Sellars quickly rejects is that unseen interiors or non-facing sides are merely believed in, just as one might believe in any other non-existent object that can be thought about (e.g., a golden mountain). But such 'intentional inexistence', Sellars quickly notes, does not do justice to the phenomenology of visual experience. To use Sellars' own example, when we see a red apple, we see it *as* containing a volume of white, and also *as* juicy and cool, though the whiteness, juiciness, and coolness are not seen *of* the apple; only the red facing surface is seen of the apple. As Sellars puts it,"[a]n actual coolness is bodily present in the experience as is an actual volume of white" (IKTE II.20/KTM p. 422). This, in turn, raises the further question as to how that particular red facing surface is seen as the red facing surface *of the apple*, since an apple is a physical object and not a facing surface.

Since the interior whiteness (as well as many other occurrent sensible properties) of the apple are not merely believed in, and they are not what is immediately seen of the apple, what is their mode of presentation in perceptual consciousness? Sellars' answer is that they are present in perceptual

consciousness by virtue of being *imagined*. Concepts are also at work, since the applehood of the apple is neither sensed nor imaged but conceived of. Thus Sellars remarks that "[r]oughly, imagining is an intimate blend of imaging and conceptualization, whereas perceiving is an intimate blend of sensing *and* imaging *and* conceptualization" (IKTE II.23/KTM pp. 422–423). To perceive an apple *as* an apple is to be perceptually aware, simultaneously, of what is sensed (the red facing surface), what is imaged (the white interior, the red non-facing side, the juiciness and coolness), and what is conceptualized (the apple *qua* physical object with causal and modal properties). The role of imagining here is crucial, since it supplies what Sellars thinks is not conceptualized. Though we might not introspect being aware of the interior whiteness of the apple, our awareness of it consists in a reservoir of expectations of what *would* be sensed if we were to bite into it. This expectation is not an application of concepts, since the expectation would persist even if one had no grasp of the relevant causal and modal properties of "apple" embedded in material inferences. Put otherwise, even beings that lacked the concept of *apple* would still be able to imagine the interior whiteness when visually presented with occurrent red facing surfaces.

The synthesis of sensing, imaging, and conceptualizing is performed by what Sellars, here closely following Kant, calls "the productive imagination." The role of the productive imagination in perceptual consciousness involves "the *constructing of sense-image-models of external objects*" (IKTE III.25/KTM p. 423; emphasis original). Two important features of image-models must be stressed. First, the image-model of an object is *perspectival*, whereas neither bare sense-impressions nor the concept of an object are; second, the construction of an image-model of an object necessarily involves the construction of the image-model of the perceiver's own body. There is nothing essentially perspectival to the concept of a dog, but the image-model of a dog (whether imagined or perceived) is necessarily that of a dog as seen from a particular point of view indexed to the position of the perceiver's body in relation to the dog as imagined.[4] Since "the construction is a unified process guided by a combination of sensory input on the one hand and background beliefs, memories and expectations on the other" (IKTE III.25/KTM p. 423), we should say that the role of sensory input—whether stimulation on the retina, as in visual consciousness, or the triggering of any other sensory-receptors—is to guide or constrain the construction of image-models. The image-model of the perceiver's body and the image-models of the perceptual objects *together* transform the array of sense-impressions ("stimulations of the retina") into *perspectives* on a three-dimensional object that persists in space and time.[5] This is not to say simply that we always perceive from some particular perspective, but also that we always perceive from some particular perspective *by virtue of* how the productive imagination simultaneously constructs the image-model of the perceiver's own body *together with* the image-model of the perceptual object. (Though one might be able to acquire a concept of a physical object in the

absence of proprioception or body-schema, one would not be able to *use* that concept in generating image-models.) The functioning of the productive imagination here is nicely enriched by Coates's (2007; 2009) argument that the productive imagination is not just static but *temporal*: the sense-image-models ground one's expectations, anticipations, and preparations for subsequent sensations.

In this way Sellars appears to have an adequate account of the role of embodiment in perception, and thereby the intimate relation between action and perception that has long been stressed by thinkers in the pragmatist tradition. Sellars notes as much when he says that the acquisition of knowledge of objects "also involves action in relation to these objects—if only by changing one's relative position—is a point to which he [Kant] pays less attention that it deserves. Compare C. I. Lewis's treatment of this topic in the first chapter of *An Analysis of Knowledge and Valuation*" (KTE II.13n3/ KTM p. 271). In other words, action is at work in perception by virtue of the relations between the perceiving subject and perceived object captured by the respective image-models of the perceiver's body and of the object as perceived—though it must be noted that this involvement should not be understood as a necessary or essential relationship.[6]

In his sophisticated defense of critical realism, Coates (2005) argues at length that perception involves both nonconceptual, purely phenomenal episodes—states of the perceiver's consciousness—and concepts (even if very low-level) that are both *caused* by the phenomenal episodes and *refer to* the physical objects that cause the phenomenal episodes. Only this view, Coates argues, does full justice to the distinction between perceiving and thinking. It also allows us to understand the difference between veridical and non-veridical perception. In veridical perception, the intentional object of the conceptual representation is the same as the physical object that causes the phenomenal episodes; in non-veridical perception, some "deviant causal chain" (e.g., a drug or illness) has caused the very same kind of phenomenal episode. A crucial component of Coates's argument is his rejection of direct realism, which he understands in terms of a *wholly non-causal* relation between perceptual takings and perceptible objects.[7]

On Coates's view, one of the chief difficulties with direct realism is that it must resort to disjunctivism in order to treat the difference between hallucinations and perceptions; in having a complex visual hallucination, one does not see but only seems to see. Coates objects to disjunctivism because it is committed to what he calls "the qualitative indistinguishability thesis": that "sameness of experience in the qualitative sense does not guarantee sameness of experience in the ontological sense" (p. 65). The disjunctivist is committed to this thesis just because she holds that a perceiver can be in same qualitative state whether perceiving veridically or non-veridically. By contrast, the critical realist holds that the qualitative and the ontological do not come apart this way; as an ontological claim, perceptual experiences supervene solely upon what is inside the skull, and the difference between

veridical and non-veridical perceptual experiences lies solely in how they are causally brought about.[8]

I conclude the discussion of the productive imagination by briefly noting a cost in Sellars' account of embodiment. Since veridical perception requires that the productive imagination be guided by sensations, the productive imagination itself functions *identically* in perception and imagination in the narrow sense (as when I dream, hallucinate, or deliberately construct a mental image). The difference is that in veridical perception, the construction of image-models (both of the object and of my own body) is guided by sensory input, and in non-veridical perception, it is not primarily guided by any worldly inputs. As a result, our very embodiment is neutral to whether perception is veridical; the image-model of my own body is constructed by my productive imagination, whether constrained by sensory input or not. The body is present in perceptual consciousness as imagined and is constrained by the inputs of senses that are curiously disembodied—at any rate, not as fully and richly embodied as they are in earlier American pragmatists (esp. Dewey) and in certain figures in the phenomenological tradition (e.g., Merleau-Ponty). Sellars' account of the productive imagination does not explicitly attend to the formation of cognitive processes required for the organism to develop the right interconnections between motor impulses and sensory receptivity, and that veridical sensory input is important for *acquiring* an image-model of one's own body in the first place.

I now want to turn to Sellars' analysis of the role of sense-impressions in perception. In keeping with the general strategy sketched by deVries—that "transcendental structures must be realized in causal structures"—I want to distinguish between, on the one hand, the project of reflecting on the necessity of external constraint on perception, and on the other hand, any specific account of the causal items that play this constraining role.[9] To identify the necessity of constraint, Sellars introduces the concept of "sheer receptivity," whereas Sellars thinks of "sense-impressions" as the corresponding causal items that play this constraining role *in rerum natura*. That is, we have a transcendental reflection on the nature of our cognitive activity, which shows that empirical content as such requires that the agent have the capacity for sentience as well as sapience. This reflection counts as "transcendental" in the weak sense that any cognitive agent, if it is to have the kinds of cognitive experiences that we have, must be able to form judgments about a world that is external to and independent of it. But in addition to this transcendental notion, Sellars also thinks that we need a causal explanation of sense-impressions or sensa as items that are introduced, via theoretical postulation, as having the right causal powers to implement the role assigned by transcendental reflection to sheer receptivity as such.

The transcendental reflection that motivates the concept of sheer receptivity is grounded in Sellars' interpretation of the Kantian distinction between "intuitions" and "concepts."

Sellars argues that in addition to intuitive conceptual representations—that is, conceptual representations of *particulars* or of *individuals*, modeled on singular demonstrative phrases—there is a different sense of "intuition" in the receptivity of the senses, which must be *non*-conceptual in order to the requisite role of "guiding" thoughts. Thus, the productive imagination (i.e., the understanding insofar as it is playing the role of guiding sensibility) produces 'this white cube' (an intuition) modeled on 'this cube is white' (a judgment). But if intuitions, in one of their roles, are already informed by the deployment of concepts, then we need an account of "receptivity proper" to explain how our beliefs and judgments are answerable to a world that we do not create, but discover. Resolving this ambiguity in Kant's concept of intuition is important not only for understanding what Kant was trying to do, but also for understanding why subsequent thinkers did not correctly understand Kant:

> Indeed, it is only if Kant distinguishes the radically non-conceptual character of sense from the conceptual character of the synthesis of apprehension in intuition . . . and accordingly, the *receptivity* of sense from the *guidedness* of intuition that he can avoid the dialectic which leads from Hegel's *Phenomenology* to nineteenth-century idealism.
>
> (Sellars 1967, p. 16)

And much more seriously:

> Kant's failure to distinguish clearly between the 'forms' of receptivity proper and the 'forms' of that which is represented by the intuitive conceptual representations which are 'guided' by receptivity—a distinction which is demanded both by the thrust of his argument, and by sound philosophy—had as its consequence that no sooner had he left the scene than these particular waters were muddied by Hegel and the Mills, and philosophy had to begin the slow climb 'back to Kant' which is still underway.
>
> (Sellars 1967, p. 29)

That is, the distinction between the receptivity of sense and the guidedness of intuitions allow us to recognize that intuitive conceptual representations are 'guided' by something else—what he calls "receptivity proper" or "sheer receptivity."[10] When I perceptually take in how things are, my productive imagination organizes perceptual experience so that I am suitably disposed to make a claim about what I perceive. But this can be the case only if my sensory receptivity to the world has just enough structure for it to guide the productive imagination, since guidance requires constraint, and constraint entails structure. That structure, Sellars concludes, must be *representational* structure: only if sheer receptivity consists in nonconceptual representations can it possibly guide or constrain the productive imagination's deployment

of concepts. To preserve the distinction between receptivity and the products of the productive imagination (which draws upon the understanding), we must distinguish between intuitive (i.e., singular) conceptual representations and radically non-conceptual representations. The latter is how Sellars construes sensations as playing a causal role "neither epistemic nor physical" (EPM §7/SPR p. 133), but as causally efficacious states of non-apperceptive consciousness that constrain the activity of the productive imagination by virtue of being nonconceptual representations.[11]

The proper status of sense-impressions in Sellars' thinking therefore turns on a methodological principle and a transcendental result. It is a methodological principle that all transcendental structures must be reflected in causal structures, and it is a transcendental result that sheer receptivity is a transcendental structure. Hence there must be some causal structure in which sheer receptivity is reflected. On this basis Sellars then argues that sense-impressions, as causally efficacious episodes of consciousness, are the right causal structures to reflect the transcendental structure of sheer receptivity. However, it is one thing to think both that there is a transcendentally justified commitment to sheer receptivity *and* to try to locate the right causal structures for implementing that role; it is quite another to think that sense-impressions are the best candidates for doing so. In the next section, I will turn to embodied cognitive science to offer a somewhat different account of perception, one that requires revising critical realism.

2. Sensorimotor Abilities as the Productive Imagination

In recent years, the major debate within cognitive science between symbolic or computational models of the mind and connectionist models of the mind has been contested by the rise of *enactivism*. The term "the enactive approach" was coined by Francisco Varela in *The Embodied Mind* (Varela, Thompson, and Rosch 1991), which contrasts with computationalism and connectionism (as traditionally understood) by focusing on cognition as an ongoing process that links animals with environments, rather than as an activity that transpires entirely or mostly within brains.[12] Accordingly, Rowlands (2010) refers to both computationalism and connectionism as "Cartesian cognitive science" (pp. 2–3), in contrast with the focus on "4E cognitive science," or mind as "embodied, embedded, enacted, and extended" (p. 3)—a contrast nicely echoed by Wheeler's (2005) distinction between 'Cartesian cognitive science' and 'Heideggerian cognitive science'.[13]

If enactivism promises a new research program for cognitive science, we shall need to understand why that research program ought to be preferred. At present, the strongest arguments for enactive cognitive science rely on philosophical rather than strictly scientific grounds.[14] Enactive cognitive science begins with the idea that the existential phenomenology of Heidegger or Merleau-Ponty (as well as its pragmatist cousin in Dewey) is a better description of what it is to be a cognitive agent, or a better description of

mindedness as such, than competing descriptions drawn from Descartes, Hume, or Kant. That is, enactive cognitive science takes the descriptions found in Heidegger's *Being and Time* or Merleau-Ponty's *Phenomenology of Perception* as explicating the *explanandum* for cognitive science. The proper task of cognitive science, thus conceived, is to propose and test models of subpersonal cognitive machinery that causally underpin the person-level functions described by existential phenomenology.

At the subpersonal level, then, enactivist cognitive scientists are committed to two theses: that at least some subpersonal cognitive mechanisms are constituted by their role in extra-cranial functions (the *embodiment* thesis) and that at least some subpersonal cognitive mechanisms are constituted by their role in extra-somatic functions (the *embeddedness* thesis). To be a cognitive agent—to be *minded*—is to be embodied and embedded. This means that there will be at least some cognitive mechanisms that exhibit a causal 'spread' beyond the brain into the body and/or into the environment. These two theses then constrain the choice of models for thinking about the mechanisms of mindedness. Whereas computationalism treats the mind as a symbol-processing system analogous to a Turing machine, and connectionism treats the mind as a neural network, enactivism treats the mind as an embodied dynamic system (Thompson 2007, p. 4), such that "cognitive processes emerge from the nonlinear and circular causality of continuous sensorimotor interactions involving the brain, body, and environment" (Thompson 2007, pp. 10–11). We are therefore led to a new metaphor for guiding construction of testable models.

Thus, while an embodied-and-embedded cognitive system could be modeled according to functionalist principles, the requirements of embodiment and embeddedness would require a different kind of functionalist approach than those that dominated the literature in the 1970s. Likewise, although enactivists are commitment to anti-representationalism (Varela et al. 1991; Chemero 2009), in what follows I shall put this emphasis to one side; emphasizing the embodiment thesis and the embeddedness thesis does *not* require abandoning representationalism wholesale, if one were to conceive of representations as affordance-detecting and action-guiding functions of neurodynamical systems. (However, I will not comment on how some version of representationalism *could* be made compatible with Chemero's anti-representationalist account of direct perception.)

Turning to the enactivist account of perception specifically, Noë (2004) emphasizes the crucial role of *sensorimotor skills* in perceptual experience: "intuitions—patterns of stimulation—without knowledge of the sensorimotor significance of these intuitions—is blind. Crucially, the knowledge in question is practical knowledge; it is know-how. To perceive you must be possession of sensorimotor bodily skills" (p. 11). If the 'must' is restricted to what we take to be paradigmatically 'normal' cases and not inflated into logical necessity, the enactivist claims that one can learn how to perceive objects only by interacting with them over time; it is not the imagination

per se that is doing the relevant work, but the acquisition of sensorimotor skills. Contra Sellars, it is not the application of concepts and construction of images by the productive imagination that imbues sensations with cognitive significance and thereby constitutes perceptual experience; rather it is the sensorimotor skills of a living animal that play this role, crucial to which is coming to understand the way in which one's sensations co-vary or would co-vary with one's actual or possible movements (p. 15).

But what is thereby perceived, according to this style of explanation? Chemero (2009), much like Noë, grounds direct realism in the theory of how sensorimotor abilities contribute to perception. Unlike Noë, however, Chemero carefully criticizes and revises the theory of "affordances" promoted by the ecological psychologist J. J. Gibson. According to Gibson, affordances are something like the detectable properties of an environment relative to an organism. Chemero's revision—what he calls "Affordances 2.0"—is that "affordances are relations between abilities to perceive and act and features of an environment" dynamically over the time-scales of behavior and of development (pp. 150–151). Since affordances are what an animal directly perceives, an animal directly perceives the relation between its sensorimotor abilities and features of its environment. This counts as "direct," by Chemero's lights, just because the animal does not construct an inner representation of the perceptible object, as for example the visual processing theory of David Marr would have it; instead, the information necessary for suitably adjusting actions to the environment is available at the sensory receptors themselves.[15]

On this basis Noë and Chemero can give a different account of Sellars' problem of perceptual presence: how are the unobserved properties of perceived objects (e.g., interior volumes and non-facing sides) manifested in perceptual consciousness? Whereas Sellars argues that the unperceived properties are present in virtue of being imagined, Noë and Chemero can suggest that they are present by virtue of the accumulated store of implicit know-how of sensorimotor contingencies. This bodily know-how allows us to perceive the red apple as containing a volume of white that is juicy and cool, because that is how we *would* sense it if we were to cut it open or take a bite out of it. The counterfactual occurrent sensing is explained in terms of the modality of sensorimotor skills. The same skills are also drawn upon in allowing us to see *as* three-dimensional objects with which we have no first-order bodily interactions, such as clouds or perhaps the moon.

Though Sellars clearly notes the role of the body-image in constructing the relevant image-models, Noë proposes a much more intimate coordination of sensing and moving; the relevant bodily skills are *sensorimotor*, not mere sensations that have to be taken up first and then be synthesized in accordance with the concept of the perceiver's body and the concept of the object. In short, Sellars' model of the productive imagination is a "top-down" model—the concepts provide the recipes that constrain how the sensations are synthesized into image-models—whereas Noë's model is a "bottom-up" model, according to which sensorimotor skills constitute

perceptual intelligibility independently of the Understanding, assuming that the latter is construed as the capacity to apply concepts as predicates of possible judgments. The enactive approach stresses, with Merleau-Ponty, that 'to perceive is not to judge'.[16] Here it must be stressed that *perception* does not involve the logical functions of judgment, *not* that sensation does not. On Merleau-Ponty's account of perception subsequently adopted by enactivists, we should reject the idea that perception is a synthesis of sensation and conceptualization; instead it is a bodily activity with its own structures and norms distinct from those of propositionally articulated thought. Correspondingly, the task of enactive cognitive science is to propose and test models of the subpersonal mechanisms that implement embodied perceiving.

One might worry, however, as to how well justified enactivism is. Enactivism is justified, to the extent that it is, by the promise of locating subpersonal cognitive mechanisms that cohere with the person-level descriptions found in existential phenomenology. Enactive cognitive science could therefore be undermined in two different ways. First, it could be undermined if Heidegger or Merleau-Ponty are radically and fundamentally mistaken about what it is to be minded at the personal level—though I stress that it would need to be a quite radical undermining, and not simply pointing out that the phenomenology of embodiment *in their texts* fails to take into account factors such as gender, race, or disability. Likewise, enactive cognitive science could be undermined if it turned out that orthodox 'Cartesian' cognitive science yielded better explanations of subpersonal cognitive mechanisms. If epistemically virtuous scientific explanations entailed that no cognitive mechanisms were constituted by their causal spread into body or environment because all putative cases of constitution were better understood in terms of coupling, enactive cognitive science as currently formulated would have to be discarded.[17]

3. Embodied Critical Realism

We have seen that embodied cognitive science holds that an animal directly perceives affordances, understood here (following Chemero) as the relation between its sensorimotor abilities and features of its environment. The sensorimotor abilities that enable direct perception of affordances function much like the productive imagination in critical realism. According to Coates (2009), "the imagination produces in the perceiver an implicit awareness, or set of expectations, of the likely ways in which the phenomenal, or sensory, aspect of an experience will be transformed" (p. 96); in just the same way, "an animal's sensorimotor abilities manifest themselves in embodied action that causes changes in the layout of available affordances, and those affordances will change the way abilities are exercised in action" (Chemero 2009, p. 151). The question naturally arises whether the Gibsonian direct realism advocated by Noë and Chemero is preferable to the critical realism of Sellars (*peré et fils*) and Coates.

In a detailed criticism of Noë, Coates (2007) observes that Noë must help himself to a concept of phenomenal consciousness that he does not adequately explain: "unless experience can be factored into an occurrent actual component, as well as a potential component, it would be completely puzzling what the claim about 'access' to an object could amount to" (p. 102). It will not suffice to appeal simply to sensorimotor skills or abilities as a theory of perception, because that can explain at most the potential of perception, and not what is actual or occurrent to sensory consciousness. My perceptual awareness of my smartphone is partly constituted by the expectations of what I would see if I were to pick it up, or swipe the screen, and those expectations are grounded in my sensorimotor abilities that can be coupled to the smartphone's perceptible features. I do indeed directly perceive the smartphone-affordances. But Coates rightly objects that Noë takes perceptual *presence*, what is *actual* in sensory consciousness, as a primitive—as indeed does Chemero, though Chemero's explanation of sensorimotor abilities is vindicated by experiments in ecological psychology rather than human perceptual pathology.

We can, however, bring direct realism and critical realism together in what I call *embodied critical realism*. Embodied critical realism is what one gets when sensorimotor abilities take the place of the productive imagination, such that the dynamic unfolding over time of the relation between sensorimotor abilities and environmental features explains how perceptual awareness of objects is explicated in terms of expectations. When one sees a cat moving toward its food bowl, what makes that a genuine perceiving through the visual modality, and not just a visual sensation, is the implicit awareness that if one were to get up and move about, one's sensations would be expected to reliably co-vary with movement of one's body and of the objects perceived. This requires that possible movements—and the expectations, anticipations, preparations, etc., that are grounded in them—turn sensations into perceptual takings. But what is perceptually taken—the content of perceptual experience—is not the phenomenal episode *per se* but rather the affordances *qua* relations between features of the environment and sensorimotor abilities of organism.

As noted above, critical realism holds that sensations are *inner* episodes, states of sensory consciousness or of the perceiver. From the perspective of embodied critical realism, Coates's critical realist critique of direct realism conflates an insight with an error. For while Coates is right to object to Noë (and by extension Chemero) that perception requires something *occurrent* and not just *potential*, it does not follow that what is occurrent must be intrinsic, let alone intrinsic to consciousness. Relations can also be occurrent. Embodied critical realism agrees with critical realism that perception requires that there be something occurrent as well as possible (expectations, anticipations, preparations, surprisals, etc.), but reconceives these occurrent phenomena as specific actualizations of the coupling between sensorimotor abilities and environmental features. Sensations are thus relational properties rather than intrinsic properties of sensory consciousness; they are *of* but

not *in* sensory consciousness.[18] (It remains to be seen whether embodied critical realism must accept some version of disjunctivism, though *prima facie* the position implies that hallucinations do not, *contra* critical realism, share a phenomenal component with perceptions.)

Do sensations, thus reconceived, causally mediate between conceptually structured perceptual takings and physical objects, as the critical realist maintains? Recall that the critical realist holds that physical objects cause episodes in sensory consciousness, which in turn guide the deployment of concepts, via the construction of sense-image-models, and it is those concepts that then refer, in veridical perception, to the physical object that caused the phenomenal episode. Embodied critical realism accepts that there is something right about the occurrent/potential contrast but understands it in terms of occurrent and possible *relations*. The occurrent relation (obtaining over a temporal interval detectable to consciousness) functions as a causal 'anchor' that is colored by a penumbra of possibilia: the implicit awareness of possible sensorimotor actions due to the animal's sensorimotor abilities. Whereas critical realism then envisions the difference between what is occurrent and what is possible in terms of the nonconceptual, purely phenomenal states of consciousness and the imagined actions grounded in the productive imagination, embodied critical realism takes even sensations themselves to be occurrent relations that "anchor" the possible relations of further embodied action in response to features of that environment.

For example, when I perceive my cat stalking a red dot emitted by a laser, both the cat and myself are described in terms of the occurrent relation between the specific posture, shape, and orientation of our bodies at that moment and the specific layout of that environment (insofar as it is detectable by organisms with our sensory receptors). This occurrent relation then causally mediates the further deployment of our respective sensorimotor abilities in relation to that environment as she attempts to catch the moving dot that continually eludes her.[19]

What, then, of the conceptual component of perception, according to embodied critical realism? Here we face a problem familiar to readers of Sellars and McDowell: if perception is not conceptually structured at all, then it has no epistemic function, but if it has the same conceptual structures as discursively articulated thought, then it is unclear how to account for the real differences between perceiving and thinking.[20] Embodied critical realism suggests that sensorimotor abilities in sapient animals constrain the discursive activity that underpins inferentially articulated propositional contents (and hence also attitudes toward those contents). Yet we should notice that receptivity proper, in the sense that contrasts with spontaneity, requires only the Kantian view that spontaneity characterizes the Understanding, which in turn is paradigmatically bound up with judgment, and with discursively articulable concepts as predicates of possible judgment. Nothing in that conception of spontaneity requires that receptivity be utterly devoid of activity of its own distinctive kind.

Indeed it is an empiricist conviction, retained by Kant and by Sellars, that only that which is passive can count as genuinely receptive. I do not think this need not be so; on the contrary, I find myself in basic agreement with Kukla and Lance (2014) when they write, "receptivity should not be equated with passivity—my encounter with objects is receptive, not to the extent that I play a passive role in this encounter, but instead to the extent that I find or discover them to be a certain way in a direct encounter with them" (27). Indeed, nothing would be essentially lost in Sellars' contrast even if, as Noë argues, sensorimotor skills were conceptual insofar as they involve the ability to generalize across a plurality of cases, classify diverse stimuli as correlated with a single motor routine, and conversely correlate diverse motor routines with a single or small class of stimuli. We would simply require a distinction between two different dimensions of conceptual content: a dimension at work in sensorimotor skills and another at work in propositional content, whether as non-inferred content suitable for premises in inferences (what Sellars calls "language-entry transitions") or as conclusions of practical inferences (the corresponding "language-exit transitions").

Thus far I have argued only that we can push critical realism toward embodied critical realism if we were to take on board key ideas from embodied cognitive science as an alternative to mainstream cognitive science. Why, however, should Sellarsians *prefer* it? Within the problematic pursued here, one important response is that it offers guidance on an omission in Sellars' theory of perception: the genesis of the productive imagination. Although Sellars (1980) excelled in recognizing that the incoherence of the Given meant that the problem of how language is transmitted from generation to generation had to be resolved within the then-new explanatory framework of behaviorism, he did not recognize that the same is true for the productive imagination itself. Instead, by restricting himself to explicating the productive imagination within roughly Kantian terms, Sellars deprived himself of recognizing the need for a parallel account of how the developing organism acquires the capacity to construct image-models in accord with an image-model of the perceiver's own body.

Enactive cognitive science offers a solution to this problem in two respects. First, by taking the explication of what it is to be a cognitive agent from Heidegger, Merleau-Ponty, and Dewey, rather than from Kant, enactive cognitive science begins a description of mindedness as always embodied and embedded. Second, by focusing on the role of sensorimotor skills as always in relation with environmental features, enactive cognitive science offers an account of how the productive imagination is brought forth: it comes into being as sensorimotor skills are gradually acquired over the course of early cognitive development. The developing organism learns which actual and possible movements are correlated with which actual and possible sensations as it navigates its environment through trial and error. Though this suggestion is at best a promissory note, without the explanatory resources of enactivism, there is a real danger than the productive imagination will be

treated as Given, in the pernicious sense that Sellars enjoins us to avoid. In that regard enactive cognitive science should attract the attention of contemporary Sellarsian philosophers, whether or not they find embodied critical realism a proposal deserving of serious consideration—though of course that is to be hoped as well.[21]

Notes

1 I do not mean to dismiss the possibility that a perceptual capacity that was selected for its role in detecting motivationally relevant stimuli can also be used to detect motivationally irrelevant stimuli. Likewise, a sentient animal born without any motor capacities might be able to discriminate between stimuli, though to a severely attenuated degree.

2 For this understanding of 'transcendental', and esp. of Sellars as a transcendental philosopher in this sense, see Westphal (2010).

3 The direct realist need not deny that there are causal intermediaries at the subpersonal level of cognitive machinery; she denies only that a philosophical account of perceptual experience *at the personal level* requires epistemic or causal intermediaries.

4 Though sensations may encode perspectival data, our understanding of that encoding is part of our scientific theory of sensations. Sellars' concern at this point in the dialectic is phenomenological, not explanatory—precisely in order to reveal the limits of phenomenological description.

5 If the only perceptible objects were such that no one could physically interact with them, the productive imagination would function completely differently—if at all.

6 Hanna and Maiese (2009) argue on transcendental grounds that perception involves embodiment metaphysically necessarily, but not logically necessarily. Nothing in the present essay depends on this strong modal claim.

7 The direct realist need not deny that there are causal intermediaries in the subpersonal cognitive machinery modeled by cognitive scientists of perception; she holds only that causal intermediaries play no role in what is explicated from within the standpoint of experience.

8 Rockwell (2005, pp. 69–81) argues for embodied cognition by criticizing the assumption that we can clearly distinguish between supervenience and causation—an assumption on which Coates relies.

9 Though this is close to McDowell's (1999) distinction between constitutive and enabling explanations, McDowell does not seem to think that cognitive science is relevant to epistemology or philosophy of mind; as will be made clear, that is not my view.

10 Cf. "The pattern of Kant's thought stands out far more clearly if we interpret him as clear about the difference between *general* conceptual representings (sortals and attributive), on the one hand, and on the other, *intuition* as a special class of *non-general* conceptual representings, but add to this interpretation that he was *not* clear about the difference between intuitions in this sense and sensations" (KTE II.15/KTM p. 272).

11 See Schellenberg (2006) for a sophisticated justification of this interpretation.

12 I say "as traditionally understood" because more recent work in the philosophy of cognitive science undermines the contrast between computational approaches and embodiment approaches; see esp. Clark (2016).

13 The enactive approach is further developed in Chemero (2009), Noë (2004), Stewart et al. (2014), and Thompson (2007) and with similar approaches taken in Clark (1997; 2016), Rockwell (2005), and Wheeler (2005).

14 Gallagher (2015) argues, following a suggestion by Cecilia Heyes, that enactivism is more of a 'philosophy of nature' than 'scientific research program': "On the one hand, enactivism makes empirical claims . . . and in this sense it resembles a research program that can suggest new experiments and new ways of interpreting data. On the other hand, its emphasis on holism [e.g. brain-body-environment dynamics—CS] presents problems for empirical investigations. One does not get far in experimental science without controlling for variables. With respect to its holistic approach, enactivism resembles a philosophy of nature" (p. 295).
15 See Rowlands (2010) for a comparison of the 'internalism' of Marr and the 'extenalism' of Gibson in their respective theories of perception (pp. 26–30; 33–37).
16 Cf. Merleau-Ponty: "to perceive in the full sense of the word (as the antithesis of imagining) is not to judge, but rather to grasp, prior to all judgment, a sense immanent in the sensible" (2012), p. 36.
17 The "coupling-constitution fallacy" takes it that it is a fallacy to infer, from "x is coupled to y", that "x and y constitute z"; see Adams and Aizawa (2001); for a critical response, see Rowlands (2010).
18 Neurobiologically, sensations are encoded egocentric spatio-temporal information about actual features of one's environment over time. On Clark's (2016) intriguing "predictive processing" model, we can think of sensations as encoding information so as to enable prediction errors to be transmitted to cognitive maps 'upstream'. The relation between predictive processing models and sensorimotor contingency theory lies outside the scope of this paper.
19 Unbeknownst to her, her inability to do so is due to my subtle hand gestures as well as the physics of light, but this is a distinction in our respective conceptual powers rather than in our respective perceptual powers *per se*.
20 For the latter concern, see esp. Baz (2003).
21 I would like to thank David Pereplyotchik for his extensive criticisms on a previous draft. I have benefitted greatly from having been pressed on numerous points. I would also like to thank Deborah R. Barnbaum for her role as conference organizer and co-editor of the present volume. Thanks are also due to the participants in the Sellars in a New Generation conference. An earlier draft was also presented to Sellars' Legacy: Consequences, Ramifications, New Directions at the American University of Beirut; my deepest thanks to the organizers and participants at that conference as well.

References and Further Reading

Adams, Frederick and Kenneth Aizawa. "The Bounds of Cognition" *Philosophical Psychology* 14 (2008): 43–64.
Baz, Aver. "On When Words Are Called For: Cavell, McDowell, and the Wording of the World" *Inquiry* 46, no. 4 (2003): 473–500.
Chemero, Anthony. *Radical Embodied Cognitive Science*. Cambridge: The MIT Press, 2009.
Clark, Andy. *Being There: Putting Brain, Body, and World Together Again*. Cambridge: The MIT Press, 1997.
Clark, Andy. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. New York: Oxford University Press, 2016.
Coates, Paul. *The Metaphysics of Perception: Wilfrid Sellars, Perceptual Consciousness and Critical Realism*. New York, NY: Routledge, 2007.
Coates, Paul. "Perception, Imagination, and Demonstrative Reference: A Sellarsian Account" in *Empiricism, Perceptual Knowledge, Normativity, and Realism*, edited by Willem deVries, 63–100. New York: Oxford University Press, 2009.

DeVries, Willem. "Sellars vs. McDowell on the Structure of Sensory Consciousness" *Diametros* 27 (2011): 47–63.

Gallagher, Shaun. "Do We (or Our Brains) Actively Represent or Enactively Engage with the World?" in *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science*, edited by Andreas Engel, Karl Friston, and Danica Kragic, 285–296. Cambridge: The MIT Press, 2015.

Hanna, Robert and Michelle Maiese. *Embodied Minds in Action*. New York: Oxford University Press, 2009.

Kukla, Rebecca and Mark Lance. "Receptivity and Intersubjective Experience" *The Southern Journal of Philosophy* 52, no. 1 (2014): 22–42.

Levine, Steven. "Sellars and Nonconceptual Content" *European Journal of Philosophy* 23, no. 4 (2016): 1–24.

McDowell, John. "The Content of Perceptual Experience" in *Mind, Value, and Reality*, 341–358. Cambridge: Harvard University Press, 1999.

Merleau-Ponty, Maurice. *Phenomenology of Perception*. Translated by Donald Landes. New York: Routledge, 2012.

Noë, Alva. *Action in Perception*. Cambridge: The MIT Press, 2004.

O'Shea, James. "Conceptual Thinking and Nonconceptual Content: A Sellarsian Divide" in *Self, Language and World*, edited by James O'Shea and Eric Rubinstein, 205–228. Atascadero: Ridgeview, 2010.

Rockwell, W. Teed. *Neither Brain Nor Ghost: A Nondualist Alternative to the Mind-Brain Identity Theory*. Cambridge: The MIT Press, 2005.

Rowlands, Mark. *The New Science of the Mind: From Extended Mind to Embodied Phenomenology.* Cambridge: The MIT Press, 2010.

Schellenberg, Susanna. "Sellarsian Perspectives on Perception and Non-Conceptual Content" in *The Self-Correcting Enterprise*, edited by Mark Wolf and Mark Lance, 173–196. New York: Rodopi, 2006.

Sellars, Wilfrid. "Empiricism and the Philosophy of Mind" in *Science, Perception and Reality*, edited by Wilfrid Sellars, 127–196. London: Routledge and Kegan Paul.

Sellars, Wilfrid. *Science and Metaphysics*. Atascadero: Ridgeview, 1967.

Sellars, Wilfrid. "Behaviorism, Language, and Meaning" *Pacific Philosophical Quarterly* 61 (1980): 3–30.

Sellars, Wilfrid. "Some Remarks on Kant's Theory of Experience" in *Kant's Transcendental Metaphysics*, edited by Jeffrey Sicha, 269–282. Atascadero: Ridgeview Press, 2002a.

Sellars, Wilfrid. "The Role of the Imagination in Kant's Theory of Experience" in *Kant's Transcendental Metaphysics*, edited by Jeffrey Sicha, 419–430. Atascadero: Ridgeview Press, 2002b.

Stewart, John, Oliver Gapenne, and Ezequiel Di Paolo. *Enaction: Toward a New Paradigm for Cognitive Science*. Cambridge: The MIT Press, 2014.

Thompson, Evan. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge: Harvard University Press, 2007.

Varela, Francisco, Evan Thompson, and Eleanor Rosch. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: The MIT Press, 1991.

Westphal, Kenneth. "Kant's *Critique of Pure Reason* and Analytic Philosophy" in *The Cambridge Companion to Kant's Critique of Pure Reason*, edited by Paul Guyer, 401–430. New York: Cambridge, 2010.

Wheeler, Michael. *Reconstructing the Cognitive World*. Cambridge: The MIT Press, 2005.

Part III

Metaphysics and Epistemology

7 Wilfrid Sellars Meets Cambridge Pragmatism*

Huw Price

1. Introduction

In this piece I want to connect Sellars with some philosophers I have taken to calling "Cambridge Pragmatists." I shall note some similarities, argue that each side has something to learn from the other, and propose that there's a common lesson, very close to the surface in Sellars, that both sides should embrace explicitly.

Who are these Cambridge Pragmatists?[1] One of the most prominent is Simon Blackburn, and my use of the term goes back to 2011, the year I arrived in Cambridge as Blackburn's successor. There was an opportunity to apply for conference funding from a new university scheme. With Fraser MacBride, I came up with a proposal that seemed not only an excellent fit with our own interests, but astonishingly inclusive within recent Cambridge philosophy more generally. From my side, it connected my work not only to Blackburn (that link was obvious) and to several of our apparently disparate predecessors in the same chair, such as D. H. Mellor, Anscombe, von Wright, and Wittgenstein, but also to many other distinguished Cambridge philosophers of the past century or so—Frank Ramsey, Bernard Williams, and Edward Craig, for example.

This appealingly broad church was the view that for some interesting topics, the path to philosophical illumination lies not, as other philosophers have thought, in an enquiry into the (apparent) *subject matter* of the discourse in question, but in asking about the distinctive role of the *concepts* involved—how we come to have such concepts, what roles they play in our lives, and so on. A view of this sort is very familiar in Blackburn's work on topics such as morality and modality, for example—Blackburn now calls the approach 'expressivism' and traces it in both these cases to Hume. But it also turns up, in places, in the work of a very wide range of other Cambridge philosophers. At least arguably, for example, we find it in the work of Mellor on tense, Anscombe on the first person, Craig on knowledge, von Wright on causation, Williams on truth, as well as Wittgenstein and Ramsey, famously, on various matters.

The view in question seems appropriately called a kind of *pragmatism*. It claims to understand the concepts in question in terms of their *use*—their practical role in our lives—rather than in terms of any 'corresponding' metaphysics. So, a little cheekily, MacBride and I labelled our project 'Cambridge Pragmatism'. As we were well aware, the cheek was triple-barrelled. One could find such views outside Cambridge. Many of the Cambridge philosophers on our list would not have regarded themselves as pragmatists. And there were famous pragmatists—not necessarily in quite the same sense—associated with another Cambridge! But despite or perhaps because of these blemishes, the label served our purposes very well. We organised a highly successful conference at the end of May 2012. It was held in the Winstanley Lecture Theatre at Trinity College, a few steps from Wittgenstein's remote rooms.

For me an additional advantage of the label Cambridge Pragmatism was that it made it easy to raise a question that had interested me for a number of years, that of the relation between the self-avowed expressivism of Humeans such as Blackburn, on the one hand, and Robert Brandom, on the other. Blackburn and Brandom seemed to mean different things by the 'expressivism' (Brandom taking his inspiration from Hegel, not Hume). Yet there seemed to be obvious connections, even if very little dialogue. Moreover, Brandom linked his own expressivism to pragmatism, while Blackburn certainly counted as a Cambridge Pragmatist, in my sense. So, with Brandom himself present, our conference was able to enquire into the relationship between Cambridge Pragmatism and modern American pragmatism (as it descends from the original pragmatism of the faux Cambridge, so to speak).

From this starting point, Brandom's own interest in Sellars provides one natural link to the question of the relationship between Sellars and Cambridge Pragmatism. Here I'll exploit a different connection, something more like a common cause. We can link Sellars and Cambridge Pragmatists under the banner of Humean expressivism, in Blackburn's sense. I'll begin there, highlighting some similarities between Sellars on the one hand, and Ramsey and Blackburn on the other. I shall also say something about the general shape of Humean expressivism, emphasising two things: first, its deflationary consequences for metaphysics, and second, a particular kind of problem it faces—'creeping cognitivism', as I shall call it, adapting a label due to Jamie Dreier (2004).

I shall then describe Sellars' attempts to wrestle with creeping cognitivism—not under that name, but I hope it will be clear that it is the same problem. I shall identify what I take to be Sellars' solution, and propose that it is one that Cambridge Pragmatists need as well. However, the consequences are more far-reaching than Sellars or most of the Cambridge Pragmatists have realized, I think—it requires a more thoroughgoing expressivism, in a sense I'll explain. Finally, I'll raise the question whether Sellars is ready for the deflationary metaphysical consequences of this Cambridge way of developing Humean expressivism.

As I said, my conclusion will be that there are lessons to be learnt in both directions. Sellars has something important to offer to Cambridge Pragmatists in response to creeping cognitivism. But they in turn have something to offer Sellars, in their clarity about the fact that the view offers an *alternative* to metaphysics. And there's a common lesson, close to the surface but not explicit in Sellars, that both sides do well to take on board.

2. Sellars and Ramsey

Let's begin with some familiar quotations from Sellars' 'Counterfactuals, Dispositions, and Causal Modalities' (CDCM):

> We have learned the hard way that the core truth of 'emotivism' is not only compatible with, but absurd without, *ungrudging* recognition of the fact, so properly stressed (if mis-assimilated to the model of describing) by 'ethical rationalists', that ethical discourse as *ethical discourse* is a mode of rational discourse.
>
> It is my purpose to argue that the core truth of Hume's philosophy of causation is not only compatible with, but absurd without, *ungrudging* recognition of those features of causal discourse as a mode of rational discourse on which the 'metaphysical rationalists' laid such stress but also mis-assimilated to describing.
>
> (CDCM, §82)

Thus Sellars thinks that in both the ethical and causal (or modal) cases, Hume got something right. He got right what the emotivists picked up in the ethical case—the fact that, in some sense, neither ethical nor causal discourse is in the business of 'describing the world'. What Hume got wrong, in both cases, was thinking that this put these topics outside the realm of *cognitive* or *rational* discourse.

As we shall see, Sellars anticipates Blackburn on these points. However, I think that he himself is anticipated by the first and most brilliant of the Cambridge Pragmatists, Frank Ramsey. This is clearest in Ramsey's 'General Propositions and Causality' (Ramsey 1929, hereafter 'GPC'), written in September 1929, only four months before Ramsey's tragically early death. GPC begins with a discussion of the logical status of unrestricted generalizations—claims of the form '$(x)\Phi(x)$'. Ramsey argues against his own earlier view that a sentence of this form should be treated as an infinite conjunction. However, as he puts it, "if it isn't a conjunction, it isn't a proposition at all" (GPC, 134).

In other words, Ramsey's claim is that these unrestricted generalizations—variable hypotheticals, as he calls them—are not *propositional*. They are doing some other kind of linguistic job. What job? As Ramsey puts it: "Variable hypotheticals are not judgments, but rules for judging: If I meet a Φ, I shall judge it as a Ψ" (GPC, 137). Ramsey takes this to be the key to

understanding causal thinking, too—it, too, goes into the non-propositional box.

However, Ramsey spots an important difficulty for a view of this kind. If variable hypotheticals are "not judgments but rules for judging", why do we disagree about them—why do we say "yes or no to them", as Ramsey puts it? As he says, "The question arises, in what way [a rule for judging] can be right or wrong?" (GPC, 134).

Ramsey meets this challenge head-on, discussing various senses in which we can disagree with a claim of this general form. And he insists that we shouldn't be surprised by the fact that we can disagree about something that isn't a proposition. On the contrary, he claims: "Many sentences *express cognitive attitudes* without being propositions, and the difference between saying yes or no to them is not the difference between saying yes or no to a proposition" (GPC, 137, my emphasis.)

I hope that the similarity to Sellars is clear here. Sellars uses the term 'describing' where Ramsey uses 'proposition', but it is clear that they agree on two key points. First, the boundaries of the *propositional* (Ramsey) or *descriptive* (Sellars) are not where we naively take them to be—causal claims (and at least for Sellars, ethical claims) lie beyond those boundaries. Second, the boundaries of the *propositional* or *descriptive* do not line up with the boundaries of the *cognitive*. The latter category is much more inclusive. It includes causal claims (and for Sellars, ethical claims).

Ramsey also raises the question of the relationship between the kind of account of causal judgments he proposes and a traditional *metaphysics* of causation:

> What we have said is . . . apt to leave us muddled and unsatisfied as to what seems to be the main question—a question not of the *psychological analysis* but of the *metaphysics,* which is 'Is causation a reality or a fiction, and if it is a fiction, is it useful or misleading, arbitrary or indispensable?'
>
> (GPC, 141, my emphasis)

Ramsey has offered us what he calls a 'psychological analysis' of causal talk and causal generalizations—an analysis in terms of the distinctive psychological attitudes they express. Here he imagines an opponent who brushes this psychology to one side, and attempts to return to metaphysical questions: Is there is any such thing as causation? If so, what is its nature?

Ramsey doesn't respond to this opponent directly, but it seems clear that he thinks that these questions turn out to be misguided, once we understand the *psychology* of causal judgment. In a case such as this, the interesting work takes place on the side of psychological analysis. We might say that he is 'putting aside' metaphysics. As I say, Ramsey himself is not explicit about this in GPC, but we'll see that later Cambridge Pragmatists—Blackburn in particular—do make this point explicitly.

3. Humean Expressivism

I noted that Sellars took himself to be Humean in one sense but not in another. But what does an orthodox Humean expressivism involve (in this day and age, as it were)? In the original (1994) edition of his own *The Oxford Dictionary of Philosophy,* Blackburn proposed the following definition:

> Expressivism. A term used for those theories of ethical discourse that contrast ethical sentences with expressions of belief.
>
> (Blackburn 1994, 127)

However, the restriction to ethical discourse was already somewhat anachronistic. Blackburn had long been clear in his own work that the same kind of view is attractive in the modal case, and that there are close analogues with the moral case.

The current edition of Blackburn's *Dictionary* sorts this out. Its definition of expressivism begins as follows:

> Theories that take as fundamental not the thought that we always use words to describe the world, but often to express attitudes, stances, habits of inference, and so on.
>
> (Blackburn 2016, 170)

After noting that the view is well-known in ethics, the entry continues:

> Expressivism is also applied to views in other domains that stress the practical function of uses of languages rather than any function of representing facts. So there are expressive theories of causation, modality, knowledge, and truth.
>
> (2016, 170)

This is an understanding of the term that will allow us to count Sellars as an expressivist about causal modalities and counterfactual conditionals; Ramsey as an expressivist about laws, causation, and probability; and Blackburn himself—rightly stressing the parallel between the modal and the moral cases—about probability, causation, and necessity.

It is helpful to distinguish two claims normally combined in a view of this kind—I call them the *negative thesis* and the *positive thesis*. The negative thesis tells us, in what we may call *semantic* terms, what the vocabulary is *not* doing. It is not 'descriptive', 'truth-apt', 'fact-stating', 'propositional', 'representational', or something of that kind. Some expressivists—traditional noncognitivists, in particular—might use the term 'cognitive' at this point. We have seen that Ramsey and Sellars don't have that option, but they do offer versions of the negative thesis in terms of other labels on this list.

Ramsey says that variable hypotheticals are not *propositions;* Sellars that ethical talk and causal talk are not *descriptive*.

The positive thesis tells us, in what I shall call *pragmatic* terms, what the vocabulary in question *is* doing—for example, that it is expressing evaluative attitudes, or dispositions to follow a rule. An important question at this point is whether the notion of *expressing an attitude* is the only kind of pragmatic function that might be feature in this positive thesis. The question is partly terminological, because of course we could choose to define 'expressivism' in these terms. But then we would need another term if we encountered a vocabulary about which we wanted to maintain that while it is not expressive in this narrow sense, nor is it descriptive—it has some non-semantic function other than expression of an attitude. (Arguably, the disquotational theory of truth provides a suitable example.) So I prefer to use the term 'expressivism' broadly, for any view whose positive account of the function of a vocabulary lies on this pragmatic side.

As I said, contemporary versions of Humean expressivism normally combine both theses, positive and negative. A further question, to which I shall return, is whether both theses are essential to a view of this kind. I shall answer 'no'. On the contrary, I think there are very good reasons (very close to the surface in Sellars) for abandoning the negative thesis, while remaining an expressivist in the positive sense.

4. Quasi-Realism

I noted that Ramsey anticipates Sellars in insisting on the cognitive character of his non-propositional "rules for judging", and that he raises the question as to how this cognitive character is to be explained—"why we say yes or no" to variable hypotheticals, as he puts it. The Cambridge Pragmatist who has been most explicit about this explanatory project is Simon Blackburn. It is the core of what Blackburn calls 'quasi-realism'—the project of explaining the cognitive character of a vocabulary, given the expressive or pragmatic starting point.

Here is Blackburn's own definition of quasi-realism, again from *The Oxford Dictionary of Philosophy*:

> [A] position holding that an expressivist . . . account of various domains can explain and make legitimate sense of the realist-sounding discourse within which we promote and debate views in those domains. A prime application is in ethics, although there are many others. *This is in opposition to writers who think that if expressivism is correct then our ordinary ways of thinking in terms of a moral truth, for example, or of knowledge, or the independence of ethical facts from our subjective sentiments, must all be in error, reflecting a mistaken realist metaphysics*. The quasi-realist seeks to earn our right to talk in these terms on the slender . . . expressivist basis.
>
> (2016, 397, my emphasis)

In the sentence I have highlighted here, Blackburn makes the important point that quasi-realism *opposes* a certain kind of anti-realism. It rejects the metaphysical view that there are no moral facts or moral properties.

Blackburn is clear about the rejection of this kind of anti-realism at a number of places. One particularly forceful version occurs in an appendix to *Ruling Passions* (Blackburn 1998), containing Blackburn's responses to a list of commonly encountered questions. Question 18 asks: "Aren't you really trying to defend our right to talk 'as if' these were moral truths, although in your view *there aren't any, really*?" Blackburn's answer is emphatic:

> No, no, no. I do not say that we can talk as if kicking dogs were wrong, when "really" it isn't wrong. I say that it is wrong (so it is true that it is wrong, so it is really true that it is wrong, so this is an example of a moral truth, so there are moral truths).
>
> This misinterpretation is curiously common. Anyone believing it must believe themselves to have some more robust, metaphysically heavyweight conception of what it would be for there to be moral truths REALLY, and compared with this genuine article, I only have us talking as if there are moral truths REALLY. I deny that there is any such coherent conception.
>
> (Blackburn 1998, 319)

This example is from the late 1990s, but Blackburn had made the same point much earlier. In his classic (1986) paper, 'Morals and Modals', he puts it like this:

> What then is the mistake in describing such a philosophy [quasi-realism] as holding that 'we talk as if there are necessities when really there are none'? It is the failure to notice that the quasi-realist need allow no sense to what follows the 'as if' *except* one in which it is true. And conversely he need allow no sense to the contrasting proposition in which it in turn is true.
>
> (Blackburn 1986, 57)

Again, the idea is that quasi-realism *deflates* the metaphysical language. The quasi-realist is a realist in these deflated terms, and denies that there are any other terms available—in particular, that there are any meaningful terms in which he might properly be said to be an anti-realist.

As I have noted elsewhere (Price 2009), a good ally at this point is the Carnap of 'Empiricism, Semantics and Ontology', who also favours this rejection of the traditional metaphysical issue of realism versus anti-realism:

> Influenced by ideas of Ludwig Wittgenstein, the [Vienna] Circle rejected both the thesis of the reality of the external world and the thesis of its irreality as pseudo-statements; the same was the case for both the thesis of the reality of universals . . . and the nominalistic thesis that they are

> not real and that their alleged names are not names of anything but merely *flatus vocis*.
>
> (Carnap 1950, 215)

Another ally will be the Ramsey of GPC, who, as we saw, simply sets aside the metaphysical questions.

I take this 'setting aside' of metaphysics to be characteristic of Humean expressivism, and I shall return to it below. But first, returning to my quotations from Blackburn above (especially the first of them), let's note the terms in which Blackburn puts this point. He says that the quasi-realist's project is to show us that talk of *truth* and *knowledge* is perfectly in order, in the domain in question. It seems reasonable, then, to count Blackburn as someone who agrees with Ramsey and Sellars that expressivism is not in tension with the view that the domain in question is *cognitive* in character. On the contrary, we can think of the quasi-realist's project, like Ramsey's project, as being that of *explaining* this cognitive character—explaining how and why we can disagree about expressions of attitudes, for example.

One final, important point of agreement (or apparent agreement[2]) between Ramsey, Sellars, and Blackburn: They are all *local* expressivists. That is, they all hold that some statements *are* genuine propositions, as Ramsey puts it, or are genuinely descriptive, as Sellars and Blackburn put it. In other words they are all committed to a Bifurcation Thesis, as Robert Kraut (1990) has called it—to the view that there is a significant boundary of this kind, somewhere within the class of declarative claims. This brings me to an important challenge to this aspect of their common view—a challenge other expressivists also face, but which, as I'll explain, is particularly acute for Ramsey, Sellars, and Blackburn.

5. Creeping Cognitivism

We noted that Humean expressivists typically espouse a negative thesis, saying that the vocabulary in question lies on the far side of a line (the 'bifurcation') that separates claims that are genuinely propositional (Ramsey) or descriptive (Sellars) from claims that are not. The bifurcation itself is characterised in what I called semantic terms. Yet, as Jamie Dreier (2004) noted, modern versions of this position tend to be deflationist about the relevant semantic notions. It then becomes hard to maintain that there is any interesting bifurcation left, and any sense in which the negative thesis is true, of the vocabularies in question. As Dreier complains, expressivists end up sounding just like realists. It is not clear that this is bad news for expressivists—after all, it is just what the quasi-realist wanted. But it does suggest that the standard version of Humean expressivism, committed to affirming both the negative and positive theses, is hard to maintain.

Dreier calls this the problem of 'creeping minimalism'. I want to isolate a subspecies of the problem, one that seems particularly acute for Ramsey,

Sellars, and Blackburn. It is internally generated, a product of their (commendable) willingness to concede that the domains in question *are* cognitive. Having conceded this, they confront the question: "Well, if they're *cognitive,* what does it mean to say that they're not *propositional,* or not *descriptive*?" Clearly, the question carries a threat of a dilemma. Whatever these views propose as the mark of genuinely descriptive claims, it seems likely to be something that could be pressed into service as an account of genuinely cognitive judgements, too—thus undermining their own insistence that, as Ramsey puts it, "[m]any sentences express cognitive attitudes without being propositions."

I shall call this the problem of *creeping cognitivism.* Greatly to his credit, Sellars appreciates that there a very difficult problem in this vicinity. Let's see how he wrestles with it.

6. Sellars on Describing

As I noted at the beginning, Sellars thinks that Humean insights about certain vocabularies (e.g., moral and modal discourse) are compatible with the observation that these vocabularies are also rational discourses. He says, for example:

> It is just as proper to say of statements of the form, 'Jones ought to do A' that they are true, as it is to say this of mathematical, geographical, or semantical statements.
>
> (ITM, 531)

Here he is emphasizing the fact that prescriptive or normative claims are normally called true or false. So he is not an old-fashioned non-cognitivist, who wants to deny truth values to evaluative claims. Yet he also takes for granted the Bifurcation Thesis—he wants to say that evaluative and modal discourse is not properly "assimilated to describing."

However, Sellars recognizes that it is by no means easy to say what 'describing' *means*. This is his attempt from 'Empiricism and Abstract Entities' (EAE), written in 1956:

> The concept of a descriptive term is . . . by no means intuitively clear. It is easier to specify kinds of terms which are *not* descriptive than to single out what it is that descriptive terms have in common. Thus, I think it would be generally agreed that the class of non-descriptive terms includes, bedsides logical terms in a suitably narrow sense, *prescriptive* terms, and the logical and causal modalities. . . .
>
> But what is it to describe? Must one be describing an object if one says something about it that is either true or false? Scarcely, for modals and even prescriptive statements (e.g., "Jones ought to make amends") can correctly said to be either true or false. Perhaps to describe an object

is to specify some of its properties and/or relations. Unfortunately, the terms "quality" and "relations" raise parallel difficulties. Is it absurd to speak of goodness as a prescriptive quality?

We are back with the question, what is it to describe? In my opinion, the key to the answer is the realization that describing is internally related to *explaining*, in that sense of "explanation" that comes to full flower in scientific explanation—in short, causal explanation. A descriptive term is one which, in its basic use, properly replaces one of the variables in the dialogue schema

> What brought it about that x is Φ? That y is Ψ.

where what is requested in a causal explanation.

(EAE, 450–51)

At this point, then, Sellars is appealing to what is sometimes called an Eleatic principle. Willem deVries has recently proposed a similar criterion—see his paper in this volume. But Sellars himself soon changed his mind. He makes this clear in the Sellars-Chisholm correspondence (ITM), written the following year.

> When I have said that the semantical statements convey descriptive information, but do not *assert* it, I have not meant to imply that semantical statements *only* convey and do not assert. They make semantical assertions. Nor is "convey", as I have used it, a synonym for "evince" or "express" as emotivists have used this term. I have certainly not wished to assimilate semantical statements to ejaculations or symptoms.
>
> It might be worth saying at this point that, as I see it, it is just as proper to say of statements of the form "Jones ought to do A" that they are *true*, as it is to say this of mathematical, geographical, or semantical statements. This of course does not preclude me from calling attention to important differences in the "logics" of these statements.
>
> I quite agree, then, that it is no more a solution to our problem simply to say that semantical statements are "unique" than it would be a solution of the corresponding problems in ethics simply to say that prescriptive statements are "unique." What is needed is a painstaking exploration of statements belonging to various (*prima facie*) families, with a view to discovering *specific* similarities and differences in the ways in which they behave. . . .
>
> (ITM, 531)

At this stage, we have what looks by a Cambridge Pragmatist's lights like a commendable focus on the idea that different bits of language are doing different jobs, in some interesting sense—a sense not immediately reducible to the observation that they are simply *about* different subject matters. (We also have a commendable recognition that ethical language isn't the only

such case.) Sellars then returns to the question of the meaning of 'descriptive', and disavows his own proposal from the preceding year.

> I also agree that the term "descriptive" is of little help. Once the "journeyman" task . . . is well under way, it may be possible to give a precise meaning to this technical term (Presumably this technical use would show some measure of continuity with our ordinary use of "describe".) I made an attempt along this line in my Carnap paper, though I am not very proud of it. On the other hand, as philosophers use the term today, it means little that is definite apart from the logician's contrast of a "descriptive expression" with "logical expression" (on this use "ought" would be a descriptive term!) and the moral philosopher's contrast of "descriptive" with "prescriptive". According to both these uses "S means p" would be a descriptive statement.
>
> (ITM, 531)

Let's be clear about Sellars' situation at this point. He wants to say that in some interesting sense "S means p" is *not* a descriptive statement—he is committed to the descriptive/non-descriptive bifurcation, and to putting ascriptions of meaning on the right-hand side of it. But he hasn't yet settled on anything that he regards as an adequate account of what it *means* to be descriptive or non-descriptive. He's rejected the proposal he made in the Carnap paper the previous year, but he hasn't come up with any alternative.

Sellars comes back to the issue in *Science and Metaphysics* (SM), several years later. This is from the Preface, where he writes about what he calls "the heart of the enterprise" of the book:

> I attempt to spell out the specific differences of matter-of-factual truth. Levels of 'factual' discourse are distinguished and shown to presuppose a basic level in which conceptual items . . . 'represent' or 'picture' (in a sense carefully to be distinguished from the semantical concepts of reference and [predication]) the way things are.
>
> (SM, p. ix)

Note that Sellars is clear that to the extent that there is a category of the genuinely factual, or genuinely descriptive, it is not to be characterised in terms of *semantical* notions such as truth or reference—on the contrary, Sellars says, it must be "carefully . . . distinguished" from those notions.

How should it be characterised, if not in these terms? Later in the book, in the introduction to Chapter 7, Sellars offers us this:

> My concern will be with what might initially be called 'factual truth'. This phrase is intended to cover both the truth of propositions at the perceptual and introspective level, and the truth of those propositions

> which, though 'empirical' in the broad sense that their authority ultimately rests on perceptual experience, involve the complex techniques of concept formation and confirmation characteristic of theoretical science
>
> Since the term 'fact' is properly used as a synonym for 'truth' even its most generic sense, so that we can speak of mathematical and even ethical facts, 'factual', in the more specific sense indicated above, should be thought of as 'matter-of-factual', and as equivalent to Leibniz's technical term '*vérités de fait*'.
>
> (SM, 116)

Again, this is a gesture towards what Sellars needs, which is a distinction between the *genuinely factual,* on the one hand, and the *factual in the generic sense,* on the other (the latter but not the former including mathematical or ethical facts, for example). But so far as I can see it is only a gesture. Telling us that 'genuinely factual' is equivalent to Leibniz's technical term does little more than to alert us to the fact that the same difficulty will arise for Leibniz, too—for Leibniz certainly hasn't wrestled with the question of how to make sense of the Bifurcation Thesis.

In one relevant respect Sellars is admirably clear, however. He stresses repeatedly that there are *two kinds of truth* in the picture; or perhaps better, two *different* things that are mistakenly conflated under the name 'truth'. One is a generic, semantic notion that Sellars characterizes like this:

> [F]or a proposition to be true is for it to be assertible, where this means not *capable* of being asserted (which it must be to be a proposition at all) but *correctly* assertible, assertible, that is, in accordance with the relevant semantical rules. . . . 'True' then means *semantically* assertible (S assertible) and the varieties of truth correspond to the relevant varieties of semantical rule.
>
> (SM, 101)

This generic notion applies to all kinds of propositions: mathematical ones, moral ones, and modal ones, for example. The other notion is much more specific, applying only to the 'matter-of-factual' vocabularies. As O'Shea describes it:

> Sellars also argues that [in addition to this generic notion] there is a further 'correspondence' dimension to truth in the specific case of what he calls *basic matter-of-fact truths*. This is a descendant of Wittgenstein's 'picture theory' in the *Tractatus*: basic matter of factual propositions in some sense form *pictures*, or 'cognitive maps,' or 'representations' of how objects or events in the world are related.
>
> (O'Shea 2007, 144)

Sellars is very clear indeed that these two notions need to be kept apart. As he says, the relation of picturing is "a mode of 'correspondence' other than truth" (TC, 54). He tells us that it is important to

> grasp the difference between the *primary* concept of factual truth (truth as a correct picture) . . . and the *generic* concept of truth as S assertibility, which involves the quite different mode of correspondence . . . in terms of which the 'correspondence' statements (i.e. equivalence statement) that '2 + 2 = 4' is true ↔ 2 plus 2 = 4 is to be understood.
>
> (SM, 119)

Again: "Picturing is a complex matter of factual relation and, as such, belongs in quite a different box from the concepts of denotation and truth" (SM, 136).

In my view, the realisation that there are two quite different notions in play in this vicinity—notions easily confused for one another—is a lesson that Humean expressivists should learn from Sellars. However, I think that not even Sellars properly understands its impact. Properly understood, I think, it means the end of the Bifurcation Thesis and commits us to a 'global' pragmatism, or global expressivism.

6. From Sellars to Global Expressivism

Recall again that Humean expressivism typically combines two theses. The *negative thesis* tells us in semantic terms what a vocabulary is not doing—it is not descriptive, fact-stating, 'truth-apt', or something of the kind. The *positive thesis* tells us in pragmatic terms what the vocabulary in question is doing—it is expressing evaluative attitudes, or dispositions to follow a rule, for example. Creeping cognitivism is the problem as to how we formulate the negative thesis, if notions such as 'cognitive', 'truth', or 'reference' and now being held to have a proper place on both sides of the bifurcation (as Sellars says the term 'cognitive' does, for example, and that the generic notion of 'truth' does).

Sellars' clarity about the fact that semantic notion of truth is generic means that it is for him a very small step to the move I recommend: We should simply *abandon* the negative thesis, and with it the idea that there is any well-grounded *semantic* bifurcation in the first place. On the contrary: all the vocabularies in question are *equally* fact-stating, *in this generic sense*. (That's what it means to say that these semantic notions are generic, after all.) As Sellars realises, this is quite compatible with retaining the *positive* thesis—with maintaining that there are very important *non-semantic* distinctions between the jobs done by different vocabularies.

The one piece of this view that Sellars doesn't quite have, in my view, is an explicit recognition that his 'picturing'-based notion of truth belongs on the pragmatic side—that it simply goes into the mix as one positive *pragmatic*

proposal about the role of particular vocabularies. But he can't possibly be far away from this conclusion. He himself insists that picturing "belongs in quite a different box" from the semantic notions. But as I noted above, the 'pragmatic' box of the positive thesis is best defined simply in opposition to the semantic box. Sellars is clear that picturing doesn't go into the semantic box, so he counts as a global expressivist by definition, in my terms. Only a lingering temptation to let 'picturing' creep back across the line to the semantic side, and so ground a semantic bifurcation after all, makes this classification seem controversial. And Sellars himself repeatedly insists that we should resist that temptation.

If we view the landscape in these terms, then we have given up entirely on the Bifurcation Thesis. In other words, we have simply abandoned the notion that there is a useful distinction to be drawn between descriptive and non-descriptive uses of language. Rather, we can find all the plurality we need at the underlying *pragmatic* level. Because we haven't abandoned the positive claims that expressivism makes at that level, our view surely counts as a form of expressivism. Yet it cannot be local, in the old sense, for we have embraced the challenge of creeping cognitivism, and recognised that there is no interesting semantic bifurcation. Accordingly, we should think of it as *global expressivism*, or *global pragmatism*. As I say, I think that Sellars himself leads us to this point, even if he doesn't quite appreciate the shape of the landscape that comes into view.

7. Two Notions of Representation

In recent work (Price 2011, Ch 1; Price *et al.*, 2013, Ch 2) I have drawn a distinction that I now take to have much in common with Sellars' separation of the generic, semantic truth, on one hand, and the 'picturing' notion of truth, on the other. My focus was somewhat broader than Sellars', concerned not with truth specifically but with the taxonomy of various notions of representation in play in contemporary philosophy and cognitive science. I proposed that it is helpful to distinguish two broad clusters of notions, and to recognise the theoretical advantages of insisting that they are distinct—of resisting the temptation to force them into the same box, to use Sellars' metaphor.

In one of my two boxes—the *e-representational* cluster, as I call it—the defining feature is environment tracking, causal covariation, indicator relations, or something of that kind. In this cluster, at least at first pass, we put the internal states that frogs use to keep track of flies, the states of thermometers that keep track of temperature, and the like. In the other box—the *i-representational* cluster—the defining feature is a role is some sort of functional, computational, or inferential network; for example, being a move in Brandom's game of giving and asking for reasons.

I proposed that notions from both of these clusters are useful for various theoretical purposes, but that we should avoid confusing them. No doubt

there are confusions we should avoid within each cluster, too, but it was the cluster-to-cluster confusion that seemed to me to be especially interesting. Indeed, in confusing the i-representational notion of propositional content with e-representational notions of word–natural world correspondence, it is the core mistake of much contemporary representationalism, in my view.

Against this background, it seems natural to regard Sellars' 'picturing' as an e-representational notion and his generic notion of truth (S-assertibility) as an i-representational notion. For Sellars, as for me, keeping these notions distinct is the key to the project of combining the insights of Humean expressivism with the recognition that moral and modal claims are full blown truth-evaluable assertions. But for Sellars, too, I think the upshot has to be a kind of global expressivism—that is, a global anti-(traditional) representationalism that comes from recognizing that propositional content and word–natural world correspondence live quite different theoretical lives.

Note that Sellars himself tends to reserve the term 'representation' for 'picturing'—what I call e-representation. In one sense, this is merely a terminological preference. (Other distinguished pragmatists have insisted to me that only i-representation deserves to be called representation.) But it may make it harder to appreciate that Sellars is something very close to a global expressivist, or global pragmatist. Pragmatists are traditional enemies of representationalism, after all. But Sellars is not a representationalist in the sense in question—the sense best characterized, as above, as a *confusion* of e-representation with i-representation. Sellars avoids that confusion by insisting that 'picturing' and S-assertibility belong in different theoretical categories. With that sorted out, pragmatism has what it needs, and nothing hangs on how we choose to label the two boxes in question.

7. Is Sellars Ready to Set Aside Metaphysics?

Finally, to the lesson I propose that Sellars might learn from Cambridge Pragmatists, and from contemporary Humean expressivists in general. I noted earlier that one of the characteristics of Humean expressivism, explicit in Ramsey and Blackburn, is a 'setting aside' of metaphysics, and a deflation of traditional metaphysical issues of realism and antirealism. I want to finish by proposing that Sellars needs to go this way, too.

In recent work (Price *et al.*, 2013, Ch 3) I have proposed that alongside the distinction between e-representational and i-representational notions in play in contemporary philosophy, we need to recognise a corresponding distinction between two notions of *world*. One notion (the 'e-world', as I called it) is the natural world, the object of study of science in a broad sense. The other notion (the 'i-world') is something like 'all the facts'—everything we take to be the case. As in the case of e-representation and i-representation, both notions here are to some extent clusters, capable of being refined in various ways. But the important thing is to recognise that they are distinct, and answerable to different considerations. We look to science for answers

to questions about the e-world, but often to different domains of enquiry altogether for questions about the i-world. The i-world is equally at home with mathematical and moral facts, for example.

Again, this distinction might remind us of Sellars. In one of the passages quoted above, Sellars notes that "the term 'fact' is properly used as a synonym for 'truth' even its most generic sense, so that we can speak of mathematical and even ethical facts" (SM, 116). Sellars makes this point to distinguish this generic use of 'fact' from "'factual', in the more specific sense . . . thought of as 'matter-of-factual', and as equivalent to Leibniz's technical term '*vérités de fait*'" (SM, 116).

For me, the distinction between i-world and e-world is intended to further the expressivist project of setting aside many of the concerns of contemporary metaphysics. In this case, I have in mind the kind of metaphysical naturalism that maintains that the natural world is 'all there is' (i.e., that declares itself to be 'realist' about the natural world and 'antirealist' about anything else). I want to say that this view is trivially true or trivially false, depending on whether we mean the e-world or the i-world when we talk about 'what there is'. In neither case is there an interesting philosophical issue—the appearance that there is one rests on confusing these two senses of 'world'.

My closing suggestion is that Sellars needs to go this way, too. That is, he needs to accept for *facts,* as he affirms for *expressions,* that, as he puts it:

> Once the tautology 'The world is described by descriptive concepts' is freed from the idea that the business of all non-logical concepts is to describe, the way is clear to an *ungrudging* recognition that many [**facts**] which empiricists have relegated to second-class citizenship . . . are not *inferior, just different*?
>
> (CDCM, §79, emphasis in bold mine—Sellars says 'expressions' at this point)

In other words, I think that Sellars should accept that mathematical facts, moral facts, modal facts, and the like, are "not inferior, just different".

This may seem in tension with Sellars' naturalism, but much will depend on whether Sellars is really the kind of metaphysical naturalist mentioned above, who thinks of naturalism as an *ontological* doctrine. Elsewhere (Price 2004) I have contrasted that kind of naturalism ('object naturalism', as I called it) to what I termed 'subject naturalism'—a philosophy that begins with the recognition that we humans are creatures of the natural world, and seeks to make sense of our thought and talk on that basis. This is the naturalism of Hume, among many others, in other words. If Sellars is content to be a subject naturalist, then he will have no difficulty at all in accepting this further deflationary consequence of Humean expressivism.

8. Conclusion

To sum up, I have argued that in his emphatic distinction between two notions of truth, Sellars has a basis for the response that Humean expressivism needs to creeping cognitivism. It involves an explicit rejection of the semantic Bifurcation Thesis. The result is in an important sense an anti-representationalist position, because it gives up a link at the core of orthodox representationalism between propositional content and word–natural-world correspondence. In Sellars' terms, the former keeps company with S-assertibility, the latter with 'picturing', and these notions simply live in different boxes. In effect, then, Sellars is already a global expressivist, and Cambridge Pragmatists should follow him down that path.[3] But Sellars in turn needs the metaphysical quietism that Cambridge had with the Ramsey of GPC, and later with Blackburn. And his naturalism should be that of Hume, not the object naturalism of much of contemporary metaphysics.

Notes

* This piece is based on a talk presented at *Sellars in a New Generation,* Kent State University, 2015. I am very grateful to Deborah R. Barnbaum and David Pereplyotchik for their invitation to participate in this conference, and for their assistance afterwards in arranging a transcript of my talk. I am also grateful to Lionel Shapiro for many insights into Sellars' views on representation.

1 I draw here on my account in Price (2017).

2 Cheryl Misak (2017) maintains that by 1929 Ramsey had rejected enough of the Tractarian picture of language to deserve to be counted as a 'global' expressivist, in my terms—in other words, that despite the contrast Ramsey draws in GPC between propositions and other claims, his real view by that point is that there are no propositions at all, in the Tractarian sense. I have some reservations about Misak's claim, though I agree that the Ramsey of GPC could not have been far from the global view—see Price (2017) for discussion. For present purposes, for expository convenience, I'll take for granted that Ramsey holds the 'local' view apparently on offer in GPC.

Note also that Blackburn (2017) now expresses doubts about local expressivism. Again, I'll ignore this recent development for expository convenience.

3 Perhaps Ramsey is there already—see fn 2.

References and Further Reading

Blackburn, S. 1986. 'Morals and Modals.' In *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth and Logic*, edited by G. MacDonald and C. Wright. Oxford: Blackwell, 119–141. Reprinted in Blackburn (1993), 52–74. Page references here are to the latter version.

Blackburn, S. 1993. *Essays in Quasi-Realism.* New York: Oxford University Press.

Blackburn, S. 1994. *The Oxford Dictionary of Philosophy.* Oxford: Oxford University Press.

Blackburn, S. 1998. *Ruling Passions: A Theory of Practical Reasoning.* Oxford: Oxford University Press.

Blackburn, S. 2016. *The Oxford Dictionary of Philosophy*, 3rd edition. Oxford: Oxford University Press.

Blackburn, S. 2017. 'Pragmatism: All or Some or All and Some?' In Misak and Price (2017).

Carnap, R. 1950. 'Empiricism, Semantics and Ontology.' *Revue Internationale de Philosophie* 4:20–40. Reprinted in *Meaning and Necessity: A Study in Semantics and Modal Logic,* 2nd enlarged edition, Chicago: University of Chicago Press, 1956, 205–221. Page references here are to the latter version.

Dreier, J. 2004. 'Meta-Ethics and the Problem of Creeping Minimalism.' *Philosophical Perspectives* 18:23–44.

Kraut, R. 1990. 'Varieties of Pragmatism.' *Mind* 99:157–183.

Misak, C. 2017. 'Ramsey's 1929 Pragmatism'. In Misak and Price (2017).

Misak, C., and Price, H., eds. 2017. *The Practical Turn: Pragmatism in Britain in the Long Twentieth Century.* Oxford: Oxford University Press.

O'Shea, J. R. 2007. *Wilfrid Sellars: Naturalism with a Normative Turn*. Cambridge: Polity Press.

Price, H. 2004. 'Naturalism without Representationalism.' In *Naturalism in Question*, edited by David Macarthur and Mario de Caro. Cambridge, MA: Harvard University Press, 71–88. Reprinted in Price (2011), Ch. 9.

Price, H. 2009. 'Metaphysics after Carnap: The Ghost Who Walks?' In *Metametaphysics*, edited by David Chalmers, Ryan Wasserman and David Manley. New York: Oxford University Press, 320–346. Reprinted in Price (2011), Ch. 9.

Price, H. 2011. *Naturalism without Mirrors*. New York: Oxford University Press.

Price, H. 2017. 'Ramsey's Ubiquitous Pragmatism.' In Misak and Price (2017).

Price, H., Blackburn, S., Brandom, R., Horwich, P., and Williams, M. 2013. *Expressivism, Pragmatism and Representationalism.* Cambridge: Cambridge University Press.

Ramsey, F. P. 1929. 'General Propositions and Causality.' In *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, edited by D.H. Mellor. London: Routledge and Kegan Paul, 1978, 133–151.

8 An Incoherence in Sellars' Error Theoretical Account of Color Concepts

Kevin Fink

For Wilfrid Sellars, there are two ways to give an account of the ontological status of color: on the order of being (*ordo essendi*) and on the order of knowing (*ordo cognoscendi*). Moreover, there is an important sense in which the former account is justified and supported by conclusions arrived at in investigation of the latter.[1] Sellars concludes his intricate mythological story of our evolving knowledge of color with the claim that, in the final analysis, colors exist only at the level of sensations: we falsely project them from the mind as properties of physical objects, but physical objects possess no such properties. When we properly understand the circumstances of this projection, we may come to appreciate it as false, and though Sellars is not so clear about this, it seems that we could potentially refrain from such unwarranted projections. However, after sketching Sellars' story about color on the order of knowing, I will argue that his conclusion that we can knowingly and meaningfully experience qualities such as color (whether or not we attribute them to physical objects) despite believing that physical objects actually possess no such properties, is incoherent.

In the first three sections of the paper, I outline the beginning three stages of Sellars' mythological story of the evolution of our color concepts. While Sellars' story of the conceptual evolution of our color concepts continues further than this, I do not develop it any further. Instead, I stop at his so-called "relocation story," according to which color properties are essentially relocated from the mind-independent world, where their proper objects are physical objects, to the minds of perceivers, where their proper objects are sense impressions, or states of such perceivers. In the fourth and fifth sections of the paper I develop my criticism of Sellars' relocation story, and the error theory he takes to follow from it. I argue that, according to Sellars' own view of the requirements for conceptual contentfulness, it is only in conjunction with our holding color to be a property truly possessed by physical objects that we can even coherently deploy an analogous concept of color as a property of sense impressions.

The Story of Color on the Order of Knowing: (1) The Pre-Pre-Socratics

Sellars takes our original notion of color to be that of a stuff, and he attributes this conception to those whom he calls the "pre-pre-Socratics."[2] In his Notre Dame Lectures, he says:

> I want you to think of the objects around you as three-dimensional solid conglomerations of color, they are made of color, I want you to think of color as the very stuff of which they are made. . . . The nice thing about this pink ice cube is that if you take the example seriously you begin to think of pink as a stuff. It's a cubicle chunk, if you will, of pink. . . . So that if we look now for our object of perception proper, it looks as though a paradigm case is going to be a cubical chunk of pink, and let it be so.
>
> (WSNDL: 299–300)

Typically, Sellars favors the example of the pink ice cube as a means of illustrating the non-intentional component of perceptual experience—the way in which something is present in our perceptions of physical objects other than as "merely believed in" (SRPC: 437). The pink ice cube illustrates this because we do not merely *think* of it as being colored through and through. We do not merely *think* of it as being pink on the opposite side; rather its pinkness and cubicity are (somehow) wholly present in our perception. However, although the emblematic naïveté of the pre-pre-Socratics consists in their not yet having any notion of phenomenal experience or of sense impressions, by "object of perception proper" Sellars intends, in this passage, simply an item of conscious awareness—one that may or may not correspond to or resemble an actual, material object (WSNDL: 300).[3]

Given that the pre-pre-Socratics lack any notion of phenomenal experience or sense impressions, they are incapable of referring to or categorizing bits of their perceptual experience simply as items of conscious awareness, which may or may not correspond to or resemble actual, material objects. The concepts that they do have include those of pinkness and of cubicity—which pertain to the "objects of perception proper," but as categorized by them in a particular way, namely, as physical objects. The story Sellars tells, then, is one that begins with such objects of perception (i.e., with what is "strictly speaking perceived") (WSNDL: 300) and tracks the evolution of the concepts that we use to categorize them. Essentially, the story Sellars tells about color is put forward as a story of conceptual change.

The first major conceptual change in the story is brought on through recognition, on the part of the pre-pre-Socratics, of the fact that construing colors as having the categorial status of "stuffs" does not advance the goals of explanation and prediction and so does not help us get along in the world.[4] The result is a *revision*—not a replacement—of our concept of

color such that its categorial status is transposed from that of a substance to that of a property or mode. Colors are no longer stuffs that are proper *parts* of physical objects—they are now taken as properties or dispositions of such objects. That is the first major evolution of the concept of color on the order of knowing.[5]

The Story of Color on the Order of Knowing: (2) Genius Jones' Theory of Sense Impressions

The second major evolution of the concept of color on the order of knowing is the introduction of sense impressions. This is done by the mythical genius Jones in EPM. Jones' introduction of sense impressions is not the introduction of a language to describe occurrences of which we were already somehow consciously aware *as such*. That it is part of a theory means, in part, that sense impressions are newly posited as states of perceivers designed to play an explanatory role. Their introduction makes possible a distinction, previously unavailable in the language, between, say, *seeing a red triangle*[6] and *its merely looking to one that there is a red triangle*. This distinction, then, allows us to explain perceptual errors. However, the explanatory role of sense impressions is actually more complex than this. Sellars says, "I regard sense impressions as necessary elements in a theory adequate to explain not just 'human behavior,' but also 'perceptual propositional attitudes'" (SSIS: 400). One upshot of this is that sense impressions also explain *veridical* perception and *normal* behavior.

How are we to understand these newly posited states of perceivers? Sellars introduces them through the use of a model:

> [T]he model is the idea of a domain of 'inner replicas' which, when brought about in standard conditions, share the perceptible characteristics of their physical source. . . . [T]he model for an impression of a red triangle is a *red and triangular replica* . . .
>
> (EPM: 191)

Red and *triangular* are properties of physical objects that are the properties they are, at least in part, because of the way they are causally related to other physical properties. Sense impressions, however, cannot be, strictly speaking, red and triangular, but they will have other, analogous properties. As Sellars says,

> The essential feature of the analogy is that visual impressions stand to one another in a system of ways of resembling and differing which is structurally similar to the ways in which the colours and shapes of visible objects resemble and differ.
>
> (EPM: 193)

In SK, Sellars elaborates on this analogy:

> [T]he manners of sensing are analogous to the common and proper sensibles in that they have a common conceptual structure. Thus, the color manners of sensing form a family of incompatibles, where the incompatibles involved are to be understood in terms of the incompatibilities involved in the family of ordinary physical color attributes. And, correspondingly, the shape manners of sensing would exhibit, as do physical shapes, the abstract structures of a pure geometrical system.
>
> (SK: 313)

The analogy, then, as deVries says, "is principally a matter of the relational structure of the two property fields" (deVries, 2005: 209). DeVries worries that since the analogy only concerns extrinsic properties (i.e., "relational structures of property fields") it may fail to ensure an intrinsic characterization of sense impressions. Two considerations help to alleviate this worry: first, we are familiar with the intrinsic natures of the properties that form the basis for the analogy (i.e., color as a property of physical objects); second, Sellars seems to allow us to use this knowledge to grasp the intrinsic natures of colors as properties of sense impressions.[7] He says:

> The pinkness of a pink sensation is 'analogous' to the pinkness of a manifest pink ice cube, not by being a different quality which is in some respect analogous to pinkness (as the quality a Martian experiences in certain magnetic fields might be analogous to pink with respect to its place in a quality space), but by being the same 'content' in a different categorical 'form.'
>
> (FMPP: 73)

This is a puzzling use of the term "content," but it seems that Sellars intends to highlight some sort of similarity or sameness of a purely phenomenal or non-conceptual sort. Indeed, I suspect that Sellars thinks the pinkness in the two cases is the same "content" insofar as it is the same "object of perception proper"—it's just that that object is categorized differently in each case.

The Story of Color on the Order of Knowing: (3) Relocation

The third major development in the story of color on the order of knowing involves a revision to our concept of color that results from a clash of the manifest and scientific images. Although sense impressions are, in the first instance, postulated by Jones as part of a theory, Jones' fellows and pupils are successfully trained in making a reporting use of the language of sense impressions. Indeed, Jones' pupils and descendants may even come to have direct (i.e., subjectively immediate) and privileged access to their own sense

impressions. For these reasons, we can take sense impressions to be part of the manifest image—our common sense image of the world developed through observational and correlational techniques. That accounting for sense impressions does not require the postulation of unobservables is a further reason for taking them to be part of the manifest image.

The property of the colored objects of the manifest image that poses a problem for stereoscopically fusing the two images is their "ultimate homogeneity"—by which Sellars means not that colored objects are always the same color throughout, but that "colour expanses in the manifest world consist of regions which are themselves colour expanses, and these consist in their turn of regions which are colour expanses, and so on . . ." (PSIM: 35).

The trouble comes when we attempt to reconcile this feature of the manifest image with a scientific image whose basic objects are characterized only in terms of their mechanistic properties and hence are understood as non-colored. Sellars is committed to the idea that a successful stereoscopic image would ultimately reduce manifest image objects, properties, and events to those more basic ones of the scientific image.[8] But he holds that the key principle guiding such a reduction is the following:

> If an object is *in a strict sense* a system of objects, then every property of the object must consist in the fact that its constituents have such and such qualities and stand in such and such relations or, roughly, every property of a system of objects consists of properties of, and relations between, its constituents.
>
> (PSIM: 27)

Sellars takes it as evident that color as a property of manifest physical objects will fail to be reducible to systems of non-colored objects in the way required by this principle and this prompts Sellars to "relocate" colors to the mind. The story is often referred to as the "relocation story"[9] because, despite the fact that at this stage of the overall story of color on the order of knowing there is already a sense in which sense impressions are both colored and "in the mind," we have seen that this is so in a derivative sense. For, up to this point, Sellars holds that the fundamental context for ascriptions of color is as properties of physical objects.

The preceding irreducibility argument is sometimes dubbed Sellars' "grain argument." DeVries claims that there are either two grain arguments, or one argument with two stages; I will refer to them as stages. The first stage concludes with the claim that, "colours (and odours, etc.) are not really properties of physical objects per se" (deVries, 2005: 227). The second stage aims to establish that "the proper sensibles also cannot be modifications of the brain qua system of micro-particles" (deVries, 2005: 227).[10]

The conclusion of the first stage of the argument, that color cannot possibly be a property of physical objects, is what prompts the relocation. "Relocation," of course, is a matter of theory (i.e., a change in our way of

conceptualizing things rather than a change in the things themselves). Sellars thinks that color must have all along been a state of perceivers that was categorially confused or taken for a property of physical objects. Rosenberg explains the point this way:

> What colors, within the Manifest Image, then *are*, in the final analysis, are conscious states of perceivers—states which, however, are systematically *taken* by those perceivers to be "independent existences," color quanta that are constituent ingredients of causally interactive spatio-temporal physical objects.
>
> (Rosenberg, 1982: 325–26)

Again, Sellars has much more to say by way of working out the details of the relocation (e.g., dealing with the second stage of the grain argument), but this need not concern us here, since my objection is to the coherence of the relocation story itself.

Re-evaluating Sense Impressions: What Really Makes Our Concepts of Them Intelligible?

As Sellars presents the story, it is tempting to think that Jones' new sense impression predicates become intelligible solely on the basis of our grasping the model of inner replicas and the analogy with the relational structure of the family of physical color concepts. But this would be a mistake.

A kernel of evidence for the insufficiency can be found in PSIM when Sellars contrasts Jones' theory of thoughts with Jones' theory of sense impressions. In an oft-quoted, but less often fully appreciated passage, he says:

> Whereas both thoughts and sensations are conceived by analogy with publicly observable items, in the former case the analogy concerns the *role* and hence leaves open the possibility that thoughts are radically different *in their intrinsic character* from the verbal behaviour by analogy with which they are conceived. But in the case of sensations, the analogy concerns the quality itself.
>
> (PSIM: 35)

Most commentators focus on the positive feature of the sense impression analogy (i.e., that the analogy concerns the intrinsic character). While this is clearly crucial in understanding the difference between Jones' two theories, it is essential not to overlook the negative aspect of the sense impression analogy. Unlike the case of Jones' theory of thoughts, the analogy employed in Jones' theory of sense impressions is, by itself, insufficient for mastery of concepts of the newly introduced entities. Since Sellars holds that concepts are constituted by their functional or inferential roles (IM, MFC, Brandom 2015), it is the theory of thoughts and *not* the theory of sense impressions

that makes immediately intelligible the concepts of the entities it introduces (since, in the former case, the analogy concerns precisely the conceptual role). Although, as discussed above, the sense impression analogy concerns the relational structures of the two property fields and in this sense, concerns more than just intrinsic properties, the crux of my argument will be that such relational properties are insufficient for giving the functional or inferential role in the way necessary for the concepts to be intelligible.

Although Sellars intends the overall story of color to be one of conceptual change, Jones' theory of sense impressions is *not*, taken by itself, a story of conceptual change. That is, Jones' theory is not, primarily, about a change in Jones' color concepts, but about the *introduction* of an entirely new concept—that of sense impressions of color.[11] We understand the conceptual role of ordinary color concepts (i.e., color as a property of physical objects), but we cannot assume that an analogous conceptual role holds of sense impressions of color. Moreover, it follows merely from the stipulation that we really do have two different concepts that their respective conceptual roles must be different.

What would be required to spell out the conceptual role of the newly introduced concepts of sense impressions—indeed, what is required in spelling out the conceptual role of *any* concept—would be to spell out their contribution to the system of language-entry, language-exit, and intra-linguistic uniformities (i.e., the normative functional role they have in licensing reports, inferences, and behaviors). For Sellars, spelling out a conceptual role in such a way is, at the same time, spelling out both part of a system of semantical rules of criticism and part of a system of laws of nature.[12]

Although the existence of such laws is absolutely crucial to the intelligibility of the new sense impression predicates, that fact is not something that Sellars draws attention to in connection with the theory of sense impressions. Nonetheless, Sellars summarizes the general point in the title of an early paper—"Concepts as Involving Laws and Inconceivable without Them" (CIL)—and it is a feature of concepts that remains at the core of his thinking. In that piece and others on the topic, Sellars defends at least two claims about the minimal conditions for being a concept that are relevant here: (1) concepts are defined by laws,[13] and (2) these laws must include what Sellars calls "material invariancies" of the concept (i.e., laws that go beyond what would be deducible from purely logical or merely structural principles defining the concept, or the family to which it belongs via relations to other concepts or families of concepts). In CIL, Sellars thinks of the "material invariancies" in terms of possible worlds (which he calls "histories"). A *family* of possible histories is determined by a shared set of universals exemplified in each history of the family. Material invariancies are non-compossible exemplifications that are not logically contradictory, and as such: "these invariancies restrict the family to less than what we referred to as the 'logically possible arrays of exemplifications of the universals'—and are therefore not the invariancies which are exhibited in the formulae

of logic . . .” (CIL: 301). So, for example, given that ‘cube’ and ‘banana’ are universals in our family of possible histories, it is a law of nature (not logic) that they are not compossible of a given object.

Sellars detects a circularity in resting at the general level of merely structural relations and not appealing to the “material invariancies.” He expresses this in the following passage from CIL:

> Now it might seem that the fact that universals fall into this pattern of determinables and determinates provides us with at least a partial answer to the question, “In virtue of what are two universals different?” If two most determinate universals, Φ and Ψ, fall under different determinables, of course they must be different! Yet a moment’s reflection shows that we cannot rest here. How are we to understand the difference of two determinables? How are we to understand the fact that a most determinate universal is a specification of one determinable rather than another? Redness isn’t red; nor is Color a case of Color; Φ doesn’t exemplify Φ. We thus find it difficult to put our finger on any distinctive contents for Φ and Ψ other than their relational properties with respect to the determinate-determinable structure. We seem, therefore, to be confronted by the following paradox: Each universal belongs where it does in the determinable-determinate structure, by virtue of being the universal it is; yet each universal is the universal it is by virtue of belonging where it does in the determinable-determinate structure.
>
> (CIL: 297–98)

The problem at the root of the paradox described here is that, while the difference between two determinables, say ‘red’ and ‘yellow’, can be put in terms of their respective positions in a determinable-determinate structure (e.g., red is not yellow, yellow is not red; burgundy is a shade of red but not of yellow, canary yellow is a shade of yellow but not of red) such apparent structural asymmetries actually appear entirely isomorphic at the purely formal level. Of each determinable, we are saying of it “Φ is not Ψ,” and “α is a shade of Φ but not of Ψ.” If this is the case, then what really distinguishes the determinable-determinate structure associated with ‘red’ from that associated with ‘yellow’, can’t be merely structural. It must involve an appeal to the content of the nodes in the respective structures. But if the nodes only get their content in virtue of their position in the structure, and the structures are isomorphic, then the nodes can’t be differentiated structurally.

But a similar objection clearly applies to Sellars’ sense impression analogy. Just as, the merely “relational properties with respect to the determinate-determinable structure” are insufficient to allow us to “put our finger on any distinctive contents” for any given (universal) concepts of that structure, Sellars’ sense impression analogy, as discussed above, falls short of establishing the relevant material invariancies. The relational properties that

the analogy establishes are exclusively of a structural sort. Recall the passage from SK:

> [T]he manners of sensing are analogous to the common and proper sensibles in that they have a common conceptual structure. Thus, the color manners of sensing form a family of incompatibles, where the incompatibles involved are to be understood in terms of the incompatibilities involved in the family of ordinary physical color attributes.

The sort of incompatibilities that Sellars is referring to here, in attempting analogically to give content to concepts of sense impressions of color, are merely structural ones—precisely of the sort he deemed insufficient for giving content to physical color concepts in CIL. In "Berkeley and Descartes" (BD), Sellars describes the upshot of the sense impression analogy as entailing *only* facts such as that, "a state that is a {uniform case of red} cannot also be {a uniform case of blue}" (BD: 287). This supports the claim that Sellars does not intend to extend the analogy to physical laws or material invariancies; the analogy extends solely to the structural properties of the "family of incompatibles." As such, it is insufficient, by Sellars' own lights, for giving content to the new concepts.

To see that Sellars could not possibly understand the analogy as positing, say, *phenomenal invariancies* of sense impressions that would be constitutive of our concepts of sense impressions, one need only realize that phenomenal objects and properties (including objects of both veridical and non-veridical perception, and imagination) are not *invariant* in anything like the way that physical objects and their properties are. While it may be a *material invariance*, or physical law, that bananas are never cubical, there is no such corresponding *phenomenal invariance*: I can easily dream, hallucinate, or imagine a cubical banana. Put in Sellars' terms, phenomenal invariancies (should there be any) would not restrict a family of possible histories to less than the "logically possible arrays of exemplifications of the universals"—or, at least, would not restrict such families nearly as robustly as do material invariancies—and hence could not be constitutive of conceptual content, because they would not provide sufficient resources to escape Sellars' "paradox" from CIL.

So in order that our concepts of sense impressions be intelligible and coherent, we need to spell out their contribution to the system of language-entry, language-exit, and intra-linguistic uniformities (i.e., to spell out their conceptual roles). And we need to do so in a way that not only insists on their difference from other concepts, but shows this difference through examining the laws governing their application to particulars, where these laws emphasize the relevant material invariancies. The sense impression analogy, alone, does not do this.

When we recall that the motivation for introducing sense impressions in the first place was to explain why some cognitions are veridical and others

are non-veridical, we can appreciate that these laws must properly connect sense impressions with properties of physical objects. So, qualitatively, sense impressions have properties analogous to the properties of physical objects. But this analogy only takes us so far. It only gives us a handle on the *qualitative* aspects of sense impressions (provided that we antecedently have a handle on the qualitative aspects of the analogous properties) without developing the natural laws (or material invariancies) necessary for fully intelligible concepts. This is done by the rest of the theory. Sellars says:

> Thus, sensing a pink cube is a manner of sensing which is conceived by analogy (a transcategorial analogy!) with a pink physical cube *and* which, though normally caused by the presence of a pink and cubical transparent object in front of a normal perceiver's eyes, can also be brought about in abnormal circumstances by, say, a grey object illuminated by pink light or by a pink rhomboidal object viewed through a distorted medium, or in hallucination by, for example, a probing of a certain region of the brain with an electrode, or by the taking of an hallucinogenic drug after much talk of pink ice cubes.
>
> (SK: 317, my emphasis).

And, in the EPM passage already quoted he gives the theoretical assumptions as:

> [T]he model is the idea of a domain of 'inner replicas' which, when brought about in standard conditions, share the perceptible characteristics of their physical source.

The key, then, to understanding the theoretical laws that give content to our concepts of sense impressions is not just the analogy with properties of physical objects. It is the *causal laws* connecting them with such objects. Let red_p be a property of physical objects and let red_s be a property of sense impressions. Then, the primary such law is the following: if one is a normal perceiver in standard conditions perceiving a red_p object, then one will have a red_s sense impression. Other important corollaries follow from this, but this causal law is really doing all of the heavy lifting when it comes to supplying the laws or "invariancies" that make our concepts of sense impressions intelligible. Again, Sellars certainly does not emphasize these laws in relation to the content of concepts of properties of sense impressions. But he does make opaque references to them in places—such as when he says, "In the Manifest Image, color is not identical with, though it is conceptually tied to, causal properties pertaining to the perceptual states of normal observers in standard conditions" (SSIS: 394). Although this claim is ostensibly about color as a property of physical objects, it should be clear that the conceptual tie to perceptual states does much more to define these latter than the former, since sense impressions are the (epistemologically) derivative concept.

To appreciate the nomological tie between color as a property of physical objects and sense impressions of color, an illustration will help. Suppose it is a material invariance that the only physical color compossible with an object's being a banana and being ripe is $yellow_p$. If so, then it is a law that: For all x, if x is a banana and x is ripe, then x is $yellow_p$. But given our theory of sense impressions, just as there would only be one $color_p$ property compossible with an object being both a banana and ripe, there would correspondingly be only one $color_s$ property of sense impressions compossible with a normal perceiver perceiving such an object in standard conditions: the one analogous to $yellow_p$: $yellow_s$. So in a very neat way, the invariancies that make up the conceptual laws governing sense impressions piggy-back on their physical property analogues. But, crucially: the correct picture is not that of two nomologically separate yet isomorphic domains of invariancy—the material and the phenomenal. Instead, although physical laws can be formulated without reference to the phenomenal, the reverse is not the case: there are no purely phenomenal laws sufficient for the sort of invariancies necessary to escape Sellars' paradox (from CIL). Indeed, in his discussion of Kant's theory of experience, Sellars acknowledges this basic point. He admits that: "the only inferability there is pertaining to the occurrence of sense impressions concerns their law-like relation to the stimulation of our sense organs by material things" (KTE: 280). He does not, however, seem to appreciate the consequences of this point for his error theory of color.

Criticism: The Incoherence of the Relocation Story

Can we make sense of others or of ourselves having sense impressions of color while at the same time holding that no physical objects are colored? Are our $color_s$ concepts sufficiently well-defined to stand on their own in the absence of their $color_p$ concept analogues? Sellars' view is that the evolution of our concepts on the order of knowing brings us closer to grasping truths on the order of being. The story of the evolution of our color concepts is also the story of how we come closer to knowing the ultimate truth about what color really is. Once the story (or this stage of it) is complete, we can sometimes refer back to our beliefs at earlier stages in the story as false. For instance, we can say that although physical objects were never colored, we once falsely believe they were. However, although such an evolutionary account may, from the perspective of later stages of development, describe the commitments of earlier stages as, strictly speaking, false, the plausibility of the account requires that the beliefs it attributes to us at any given stage be, at least, largely coherent and consistent with our basic philosophical commitments. I have been arguing that the relocation stage of Sellars' evolutionary account of our color concepts does not have this feature. It is incoherent, because in the absence of their physical property analogues, our concepts of sense impressions of color cannot be given any definite content.

It is one of Sellars' basic philosophical commitments that conceptual content requires the at least in principle articulation of the concept's normative functional role. Further, we have also seen that, in the case of concepts of properties such as color, this requires, at a minimum, articulating conceptual relations that go beyond merely structural features relating to "families of incompatibles." It requires that such conceptual roles be specified, at least in part, in terms of "material invariancies." Moreover, in the case of concepts of sense impressions of color, it seems that specification in terms of such material invariancies is only possible in relation to color as a property of physical objects. The relocation story, with its removal from the manifest physical world of color as a property of physical objects, removes any possibility of finding the material invariancies needed to give content to our concepts of sense impressions of color.[14]

Notes

1 Sellars elaborates on this point in his March 9, 1964, correspondence with J. J. C. Smart, when he says, of the scientific image predicates intended to do justice to the "ultimate homogeneity" of manifest colors: "These predicates would be still mere [sic] derivative in the dimension of concept formation from the sense-quality predicates of the manifest image which apply to physical things. They would, however, have a priority in the dimension of explanation by virtue of which they would describe 'what really is.' That by methodologically derivative predicates we grasp the ontologically prior, is a modern version of Aristotle's distinction between 'priority in the order of knowing' and 'priority in the order of being'," (Sellars, 1964).

2 According to deVries, the pre-pre-Socratics are part of the mythological beginnings of a story, the later stages of which are supposed to reflect "the actual history of our concepts of the sensible qualities," (deVries, 2005: 203).

3 Specifying the actual ontological status of such items of conscious awareness is the job of the account on the order of being. I have very little to say about that part of Sellars' account, since I argue that the considerations leading to it are flawed. Essentially, though, he thinks that, with color, the "objects of perception proper" are states of perceivers that are systematically taken by such perceivers to be properties of physical objects (see SK, FMPP, SRPC, and Rosenberg, 1982: 325–26, quoted below). Hence the description of his mature view as an error theory.

4 As deVries says, "Red things are not always good to eat, nor blue things bad; some red things rot, others do not" (deVries, 2005: 204).

5 Again, Sellars' "story" here is largely mythological. He is, at this point, making no pretentions to anthropological accuracy. Sellars is interested in what *logical* constraints governing concept evolution he can discover.

6 Sellars almost always uses "seeing" as an achievement term.

7 Rosenthal (2005) develops an opposing view: "There is no reason to think that individual color properties of visual sensations resemble intrinsically the color properties of physical objects" (140). While there may be independent grounds for such a claim, I think the passage from Sellars that immediately follows in this text shows pretty clearly that this was not Sellars' view, or not always his view. And that is my focus here. Further, Rosenthal's contention runs contrary to Sellars' commitment, with respect to color, on the order of being, to a single "object of perception

proper" that is taken, or categorized, by us in different ways—sometimes as a property of physical objects; sometimes as a property of sense impressions.

8 What, exactly, is required for a successful stereoscoping of the images is a controversial matter. The only point I wish to insist upon in this context is that Sellars himself believed that it calls for 'relocating' color to the mind. On this point see (SSIS: 409) and the discussion of the relocation story below. For further helpful discussion of issues concerning the relation of the two images, see (Brandom 2015, chapter 1, reprinted in this volume) and (Christias, this volume).

9 On the use of this description, see SSIS: 409, Rosenthal (2005) and (2016), and O'Shea (2009).

10 Although the first stage is sufficient for the final major revision of our color concepts that I will consider—relocation—any assessment of Sellars' solution to the grain problem must also consider the second stage. Since my interest is not Sellars' solution to the grain problem, I do not consider the second stage. My interest is in the relocation story and the revision of our color concepts it necessitates.

11 Of course, Jones theorizes sense impressions of many things besides color. I am here simply drawing attention to the fact that, with Jones' theory, we now have two distinct items: colors, and sense impressions of colors.

12 On the relation in Sellars between "laws of nature" and "semantical rules of criticism" see (Rosenberg, 2007: 52). In what follows, I will focus specifically on the connection between concepts and laws.

13 In CIL he says: "universals and laws are correlative, same universals, same laws, different universals, different laws" (296).

14 I would like to thank Richard Manning and David Pereplyotchik for their very thorough and insightful comments on multiple earlier drafts of this paper, Deborah Barnbaum and David Pereplyotchik for the excellent job they did organizing the *Sellars in a New Generation* conference at Kent State University in May 2015 and editing the current collection, Student Government at the University of South Florida for a generous travel grant to present a previous version of this paper at the *Sellars in a New Generation* conference, Kwang Sae Lee for so passionately and insightfully introducing me to the work of Wilfrid Sellars, and my parents William A. Fink and Karen E. Murray for their essential support and encouragement of my work in philosophy.

References and Further Reading

Brandom, Robert B. *From Empiricism to Expressivism: Brandom Reads Sellars*. Cambridge, MA: Harvard University Press, 2015.

———. *Making It Explicit*. Cambridge, MA: Harvard University Press, 1994.

Christias, Dionysis. "Does Brandom's 'Kant-Sellars Thesis about Modality' Undermine Sellars' Scientific Naturalism?," this volume.

deVries, Willem A. *Wilfrid Sellars*. Montreal: McGill-Queen's University Press, 2005.

O'Shea, James. "The Central Importance of Sellars' Theory of Conceptual Change" (Paper presented at College de France, Paris, May 18–19, 2009).

Rosenberg, Jay F. "The Place of Color in the Scheme of Things: A Roadmap to Sellars' Carus Lectures", *The Monist* 65 (1982): 315–335.

Rosenberg, Jay F. *Wilfrid Sellars: Fusing the Images*. Oxford: Oxford University Press, 2007.

Rosenthal, David M. "Quality Spaces, Relocation, and Grain," in *Wilfrid Sellars and His Legacy*, James R. O'Shea (ed.), Oxford: Oxford University Press, 2016: 149–185.

———. *Consciousness and Mind*. Oxford: Clarendon Press, 2005.

Sellars, Wilfrid. *The Notre Dame Lectures* [WSNDL] (bootleg version) Pedro Amaral (ed.), (2009). Available at Andrew Chrucky's website "Problems from Wilfrid Sellars," accessed May 23, 2016, http://www.ditext.com/amaral/wsndl.pdf.

———. "Some Remarks on Kant's Theory of Experience," [KTE] Reprinted in *Kant's Transcendental Metaphysics*, Jeffrey F. Sicha (ed.), Atascadero, CA: Ridgeview Publishing Co., 2002: 269–282.

———. "Empiricism and the Philosophy of Mind," [EPM] in *Science, Perception and Reality*, Atascadero, CA: Ridgeview Publishing Co., 1991: 127–196.

———. "Philosophy and the Scientific Image of Man," [PSIM] in *Science, Perception and Reality*, Atascadero, CA: Ridgeview Publishing Co., 1991: 1–40.

———. "Foundations for a Metaphysics of Pure Process," [FMPP] *The Monist* 64 (1981): 3–90.

———. "Berkeley and Descartes: Reflections on the 'New Way of Ideas'," [BD] in *Studies in Perception: Interpretations in the History of Philosophy and Science*, P. K. Machamer & R. G. Turnbull (eds.), Columbus: Ohio State University Press, 1977: 259–311.

———. "Some Reflections on Perceptual Consciousness," [SRPC] in *Selected Studies in Phenomenology and Existential Philosophy*, R. Bruzina and B. Wilshire (eds.), The Hague, 1977: 169–185. Reprinted in *Kant's Transcendental Metaphysics*, Jeffrey Sicha (ed.), Atascadero, CA: Ridgeview Publishing Co., 2002: 431–441.

———. "The Structure of Knowledge," [SK], in *Action, Knowledge and Reality: Studies in Honor of Wilfrid Sellars*, Hector-Neri Castañeda (ed.), Indianapolis: Bobbs-Merrill, 1975: 295–347.

———. "Meaning as Functional Classification," [MFC] *Synthese* 27 (1974): 417–37.

———. "Science, Sense Impressions, and Sensa: A Reply to Cornman," [SSIS] *Review of Metaphysics* 24 (1971): 391–447.

———. Correspondence with J. J. C. Smart from March 9, 1964. Wilfrid S. Sellars Papers, University of Pittsburgh Library System. Also available at Andrew Chrucky's website "Problems from Wilfrid Sellars," accessed May 23, 2016, http://www.ditext.com/sellars/smart.html.

———. "Counterfactuals, Dispositions, and the Causal Modalities," [CDCM] in *Minnesota Studies in the Philosophy of Science, Vol. II*, Herbert Feigl, Michael Scriven, and Grover Maxwell (eds.), Minneapolis: University of Minnesota Press, 1957: 225–308.

———. "Inference and Meaning," [IM] *Mind* 62 (1953): 313–338.

———. "Concepts as Involving Laws and Inconceivable without Them," [CIL] *Philosophy of Science* 15 (1948): 287–315.

9 The Causal Articulation of Practical Reality

Willem A. deVries

This paper begins the work necessary to elaborate and defend the claim I made in my book, *Wilfrid Sellars,* that the notion of *practical reality* is a significant addition to our conceptual arsenal and deserves to be taken seriously in thinking about what it means to be a realist in a naturalistic context. Among other things, this calls for a renewed consideration of Sellars' treatment of causation and the laws of nature, and it moves us decisively toward a more thorough pragmatism.

Causation and practical reality belong together because causation and reality belong together. Sellars was fond of the Platonic dictum that to be is to have power (*Sophist* 247d–e). In class, he tossed off that reference frequently. Sellars defends scientific realism in large part because science is the best road to knowledge of the law-like or causal structure of the world, and thus to knowledge of the real. If the notion of practical reality is going to do something more than just make sense, if it is going to *illuminate* something about our world for us, it will have to be tied to causation, for this will be what validates calling it a form of *reality*. I am aware that the notion of practical reality seems in overt conflict with Sellars' scientific realism, but countenancing practical reality need not require demoting the sciences from some special role in matters of ontology. I do think the issues are more complex than even Sellars, a notorious maven of the complex, appreciated. That puts all the greater burden on those who want to exploit Sellarsian thought as a base camp from which to explore further philosophical territory. In this essay, I can get us only into the foothills from which a serious expedition into the causal articulation of practical reality would have to be launched.

I. Why Practical Reality?

Let me recap why I introduced the notion of practical reality and the problems I hoped it would help us solve. Sellars draws a now familiar distinction between the manifest image and the scientific image of humans-in-the-world. The *manifest* image is the conceptual framework in terms of which we came to understand ourselves as *persons* and in terms of which, at the present time, we ordinarily comprehend our nature and place in the world.

The *scientific* image contrasts with the manifest image insofar as it employs postulational methods in developing theories to explain the way the world wags. This image, according to Sellars, is still aborning; it is yet partial and gappy, though full of promise. But Sellars also thinks that as it matures, it will come to challenge and eventually replace the manifest image. I have argued elsewhere that the description Sellars gives us of this replacement process in PSIM goes off track.[1] It would be better to say that the descriptive and explanatory resources of the sciences will come to displace the descriptive and explanatory resources currently available in the manifest image—what Sellars calls "the descriptive ontology of everyday life" (EPM §41; in SPR: 172; in KMG: 252). This means that the *prescriptive* and *justificatory* resources available in the manifest image will remain in place; they are not subject to replacement by a new scientific vocabulary. There would thus have to be a mutual accommodation of the ontology science drives us to adopt and our conception of ourselves and the world in which we act as *agents*.

But the problem now becomes evident, for the descriptive ontology of science doesn't match up with the prescriptive ontology of agency in two different dimensions. First, Person—that is, a being with the normative status we accord to moral agents—is not a scientifically natural kind, to the best of our knowledge, so agents will not be salient objects in the scientific image.[2] Second, nor will be many of the aims and objects of our agency. The tools and utensils that facilitate our lives, the social structures and practices that organize our activities, the fashions that entrance us, the art that inspires us, the activities that constitute a meaningful, flourishing life—none of these, as far as I can see, will be salient entities or cogently describable in the scientific image as Sellars envisions it. I see no reason to believe that there will be any adequate reconstruction in the terms of pure scientific theory of many of the characteristics that make these objects and activities items of concern to us. Sellars seems to think that we could learn to live our lives in terms that have hard-core scientific credentials, using that vocabulary to formulate the individual and community intentions that form the basis of our agency. That seems to me implausible at best. Suppose we were to "abandon the framework of common sense and use only the framework of theoretical science" (SRI: 189, in PP: 354). How then would objects like voting machines or Claes Oldenburg sculptures, and activities like playing Hamlet or arguing before a court, show up for us? They are themselves deeply enmeshed in complex *normative* structures and practices. In order to describe those within our scientific framework, we'd have to describe the individual and community intentions on which such normative structures and practices are founded, but we'd have to be able to do so, again, using just the vocabulary of the framework of theoretical science. Currently, of course, the intentions that underwrite the structure of voting, of the arts, or of legal argument are framed in common sense terms. That would all have to change.

Sellars thinks he has a story about this. He does not think that we would have to reconstruct *de novo* scientific descriptions of socially constructed objects and activities. Rather, he thinks that part of the move to a thoroughly scientific framework would include ceasing to *use* such manifest image terms as 'vote', 'election', 'play', and 'act' in favor of *mentioning* them. Rather than the intention 'I shall vote', I would adopt an intention that is something like 'I shall engage in the activity called "voting".' As Sellars might display it, we'll move from intentions like

Shall[I vote]

to

Shall[I 'vote']

This move allows for the socio-linguistic constitution of objects and activities without presuming their reality as such. This would be like the difference between "I saw the witch on the moors last night" versus "I saw the person we call a 'witch' on the moors last night." In other words, we would systematically adopt an anthropological stance toward the social structures and practices that organize our world, refusing to commit ourselves to the literal truth of the ordinary first-order descriptive assertions made in the language of the manifest image.

This idea now seems muddled to me. I don't mean that one cannot adopt an anthropological stance toward the social structures and practices in which one is enmeshed. But I do not think that one could *always* adopt such a stance toward them, or that the anthropological stance could be one's *only* perspective on them.

First off, I don't see how the *use* and not the mere *mention* of the first-person pronoun or some equivalent could be avoided. My intentions will always be *my* intentions, not just the intentions called 'mine'. Who would call them 'mine', if not me? I do not think that the intention at hand could be expressed as

Shall['I''vote'].

In Sellars' own definition of his 'shall' operator, a first-person reference is built into it, precisely because it is *expressive* rather than *descriptive*.

Second, I cannot see how I could live life as a self-interested but moral agent without *committing* myself to the literal truth of the descriptions and assessments by which the world is organized for me. Not even the most po-mo of us is *that* ironic. To a pragmatist especially, the notion of an ontological commitment is not the notion of a theoretical conjecture; it involves taking a stand and fundamentally structures one's agency. Prying commitment and agency too far apart empties them both of their senses.

Third, in order for the intention

'I shall engage in the activity called "voting"'

to make sense to me, someone would still have to be *using* the term 'vote' in other contexts. I suppose we could amend the intention to account for having abandoned the framework of common sense:

'I shall engage in the activity formerly called "voting".'

Having wholeheartedly made the move into the framework of the theoretical sciences, almost everything of consequence to us would have to be thought of as 'the thing formerly called X'. Could we sensibly do that?

We *can* opportunistically avail ourselves of languages we don't command. A non-German speaker can remark that sitting around the dining room table sharing food, drink, and conversation with dear friends is the kind of thing the Germans call 'gemütlich'. Similarly, one might talk about the phenomenon scientists call the collapse of the wave packet without having much understanding of quantum theory. But as far as I can see in the situation we're now envisioning, the framework and language of the manifest image would still have to be more than familiar to the denizens of the brave new world of the scientific image. They would have to know how most of the objects and activities they engage would have been described in manifest image terms and what the consequences of such descriptions would be. They would have to know, not just the "is's" and "mostlys", but the "wouldas" and "shouldas" of the things and activities they engage. Thus, they would, effectively, still have to speak manifest. I suspect, in fact, that ontogeny would have to recapitulate phylogeny to the extent that individuals could not learn the fine-grained language of scientific theory unless they first acquired the coarser and normatively loaded language of the manifest image. (I don't have an *a priori* argument for this; it is, rather, an empirical claim, but I doubt there is an ethical way to test it.)

These points follow from an honest and thoroughgoing pragmatism. Part of pragmatism is the acknowledgment that words and concepts are *tools*, and they are, thus, naturally infected with the human interests they serve. Science develops ways of speaking and thinking that minimize the direct stamp of human interests, so that we approach ever closer to an understanding of what things *are* rather than what they *are for us*. But we cannot, and therefore should not, entirely abandon the 'for us'. We should not lose our grip on the world or our recognition of the world's grip on us.

It was reflections along these lines that led me to propose that we take seriously the concept of *practical reality* as a third concept alongside Kant's concepts of *empirical reality* and *transcendental reality*. The terminology is

not perfect: the real/ideal distinction, which, in the context of the empirical versus the transcendental, is supposed to be the distinction between the mind-independent and the mind-dependent, doesn't like doing that job when it is put together with 'practical'. 'Practical ideality' is close to 'practically ideal', which sounds like the highest of goals, against which the practically real is something of a come-down. My usage is different. My talk of the practical reality of things and activities is aimed to express belief in the truth and objectivity of the prescriptive or normative aspects of such things. Practical ideality, in contrast, would then express the *subjectivity* of the normative or prescriptive.

This is, I think, a minimal sense of 'real'. It will be familiar to readers of Brandom's recent book, *From Empiricism to Expressivism*, where he defends a modal realism that he characterizes as the conjunction of three claims:

> MR1) Some modally qualified claims are true.
> MR2) Those that are state facts.
> MR3) Some of those facts are objective, in the sense that they are independent of the activities of concept-users: they would be facts even if there never were or never had been concept-users.[3]

We can characterize practical realism by adapting these claims

> PR1) Some prescriptive claims are true.
> PR2) Those that are state facts.
> PR3) Some of those facts are objective, in the sense that they are independent of the actual activities of concept-users: they are subjunctively robust and would remain facts, even if agents did not exist.

So described, I don't think there is much question that Sellars was a practical realist. He is explicit that on his conception truth is a genus that accommodates such species as empirical truth, mathematical truth, and moral truth.[4] He is equally explicit that fact-talk is just material mode truth talk, so, necessarily, true claims state facts.[5] Finally, the final chapter of SM is a search for grounds to claim the objectivity of morality. Sellars admits it remains "incomplete," but his goal is unmistakable.

The claim that Sellars is a practical realist has got to be somewhat surprising. Doesn't it clash directly with his strongly self-professed scientific realism? My argument is that it need not do so. Part of the reason there is no conflict here is that the kind of realism in question is at best minimal.[6] The physical or scientific realism that Sellars defends is surely more robust than this. Thinking about how much more robust a realism Sellars defends in empirical matters has led me to trying to think more thoroughly about causation in Sellars' thought.

II. Cheap but Robust Realism?

The kind of realism we just looked at seems *at best* minimal. It is extremely permissive. So, consider

> UR1) Some claims about unicorns are true.
> UR2) Those that are state facts.
> UR3) Some of those facts are objective, in the sense that they are independent of the activities of concept-users: they would be facts even if there never were or never had been concept-users.

All of these principles of unicorn realism are true, for

> Unicorns are mythical creatures

is true, fact-stating, and entirely objective in the requisite sense. Yet unicorn realism is not widely endorsed in the profession. If the existence of *any* true, objective claims about something or using some qualifier is sufficient for realism, then the existence of true, objective claims of non-existence would seem to suffice to establish reality in the same breath that they deny existence.

There are two responses I can envision here. One is that there is a difference between alethic modal claims about mind-independent objects and either prescriptive claims or claims about unicorns, because in the latter two cases the existence of concept users is non-trivially involved in the truth conditions of the claims, so that the third clause of the analysis does not, in such cases, turn out to be true. If there were no concept-users, there would be no agents, and no prescriptive claims would be true or state facts. If there were no concept-users, there would not even be the concept of a unicorn, and since there are no unicorns, nothing about unicorns would be true. But what about claims of the form, "If there were concept-users, there would be agents, and those agents ought to behave rationally" and "If there were concept-users, and they had the concept of a unicorn, they would recognize that 'Unicorns have a single horn' is true." Is the latter a claim about unicorns? If the truth of modally qualified claims can commit us to modal realism and the reality of possible objects, then doesn't the truth of subjunctive statements also commit us to the realities of which they speak? Once Brandom has opened the floodgates of conceptual realism, it is hard to keep anything out.

A more sophisticated response would be something like this: the claims Brandom is focused on are not *about* modalities, they actually *use* modal qualifiers; his realism is read off the *form* and not the *content* of the claims in question. We can, of course, construct *other* claims whose content concerns the forms of the claims we are primarily interested in; such sentences would make explicit what is expressed via the form of the claims we are

immediately concerned with. But those metalinguistic claims, even if they are only very abstractly or indirectly metalinguistic, are not the datum from which he infers his modal realism.

The fundamental assumption here is that the formal features of language reflect real, structural features of the world.[7] If talk about categories makes explicit the formal features of our language and thought, then, by Brandom's lights, we should be categorial realists.[8] But I think this is not the Sellarsian move. In Sellars' view, categories are highest *conceptual* kinds, and there is no straightforward inference from the structure of our conceptual framework to the structure of the world.[9] Indeed, as I've argued elsewhere, it is precisely to avoid insisting on some form of "*semantic* government of claimings by facts" with which they "correspond" that Sellars developed his conception of the non-semantic *picturing* relation between some claimings and some objects in the world.[10]

My inclination, thus, is to suspect that Brandom is hoping to get more ontological bang out of his semantic buck than he should, especially given his professed pragmatism. If Sellars is right to think that semantic terms always perform the task of classifying the functional role of linguistic/conceptual items in a broader linguistic economy that includes agency, then no ontological conclusions follow simply from the correct application of a semantic term such as, for instance, 'true' or 'fact'. There is a long, but I think unfortunate, history of trying to establish a criterion in ontology that would be either simply syntactic (such as being a proper [or maybe a "logically proper"] name) or semantic (such as being the value of a variable of quantification), but I think all such attempts are doomed to failure. The real measure for ontology is not to be found in either syntax or semantics; it is pragmatic, and not in the thin, 'linguistic' sense of pragmatic. As I remarked in my introduction, Sellars often cited with approval the passage from the *Sophist* in which the Stranger offers a definition of being in terms of the capacity (*dunamis*) to do something to something else or to be affected by something else (*Sophist* 247d-e). What we're really committed to ontologically are the things that we count on and take account of in coping with the world, even if some of that activity is in the highly rarified context of experimentation.[11] I take my point to fit within a more thorough pragmatism.

What this means for us is that the quick and dirty argument for practical realism I offered just before cannot be taken to settle the issue. In order to make the case convincingly, we have to investigate the practical effects of the practical.

III. Causation and Laws of Nature

Serious consideration of the practical effects of the practical means looking for whether there is good reason to believe that objects and activities *in their practical guise as normatively constituted or qualified* can be properly said to be causally efficacious or explanatory. The story I tell here can only

prepare the ground for a fuller argument to the effect that practical kinds of objects and activities—that is objects and activities whose *kind* is determined to an important degree by their *practical standing*—have a perfectly legitimate claim to reality in Sellars' world. This is because, as a preliminary to that fuller argument, we need to defend the respectability of 'kind' talk in ontological contexts from Sellars' attack.

The story I'm going to (re)construct will bring us face-to-face with some (but not all) of the more heterodox positions Sellars espouses. Being heterodox is not a way of being wrong. I am far from sure that the position Sellars ends up with is coherent, but I do think that the most obvious moves that would ensure coherence are moves in the direction I am recommending.

A. From Inference to Law

Brandom's work brings to the fore Sellars' belief that a crucial constitutive determinant of the meaning of linguo-conceptual items are the inferential proprieties they are involved in, where these inferential proprieties include both the proprieties of formal inference, such as *modus ponens*, and material inferential proprieties. We commonly get two different kinds of examples of materially good inferences. One kind is the inference from a determinate to its determinable (e.g., from "X is blue" to "X is colored"). That is clearly a good inference, but equally clearly not formally valid. The other standard examples (and the ones Sellars uses in "Inference and Meaning") exploit a causal connection: "It is raining, so the streets will be wet," or "This turns litmus paper red; therefore it is an acid." Both kinds of examples provide evidence that meanings and inferences are connected, both demonstrate ways in which inferential connections serve to locate claims in a "logical space of reasons," but they are significantly different in other ways. Sellars does not, in "Meaning and Inference," say anything helpful about the differences between these examples. For some illumination we have to turn to CDCM.

Sellars says there (and Brandom agrees) that description and explanation always go together

> [A]lthough describing and explaining (predicting, retrodicting, understanding) are distinguishable, they are also, in an important sense, inseparable. It is only because the expressions in terms of which we describe objects . . . locate these objects in a space of implications, that they describe at all, rather than merely label. The descriptive and explanatory resources of language advance hand in hand. . . .[12]

This is, indeed, another way of saying that inference and meaning are intrinsically tied. But it also brings to our attention the significance of the subjunctive mood and modal language, for good explanations rarely take the form of materially valid inferences in an extensional, first-order

language, despite the common use of such inferences as models of explanation.

> To make first hand use of these [modal] expressions is to be about the business of explaining a state of affairs, or justifying an assertion.
>
> (CDCM §80: 283)

This is a significant claim. The positivist view had been that the explanatory burden is carried by the (unrestricted) *generality* of a law. And indeed, generality of some kind is always present in an explanation. So it is easy to think that something like, "If the match had been dry when you struck it, it would have ignited" simply depends on the general claim that all dry matches (perhaps *ceteris paribus*) ignite when struck. But the appropriate presence of a subjunctive in such a claim does not depend solely on generality.

Sellars thinks that the ubiquitous use of modals and subjunctives in explanatory contexts is neither an accident nor a mere decoration of language. They have an expressive role that plays an important part in the practices of explanation and justification. We need to understand that role in order to understand the practical effects of the practical.

It is not generalization as such that supports the counterfactuals or subjunctives that show up in explanatory contexts. Let's call general statements that do support counterfactuals and subjunctives 'lawlike statements'. What job do they do that is signaled by putting such a statement in the subjunctive or attaching a modal qualifier to it? Brandom would have it that the subjunctive or modality signals description of a different kind of fact from an ordinary empirical fact: they are describing a *modal fact*, and he works to construct an extended sense of 'descriptive' that would enable us to apply that term to such statements. But it is pretty clear that Sellars would not follow him in this.

> To sum up, lawlike statements are not a special case of descriptive 'all-' statements. In particular, they are not descriptive 'all-'statements which are unrestricted in scope, i.e. not localized by reference to particular places, times, or objects.
>
> (CDCM §98: 299)[13]

The point Sellars wants to make here is that lawlike statements, which generally contain a subjunctive or modal expression, are not simply engaged in the task of description, What does Sellars think is really going on when we assert lawlike statements or use a counterfactual or subjunctive in an explanation? He tells us,

> It is therefore important to realize that the presence in the object language of the causal modalities (*and* of the logical modalities *and* of the deontic modalities) serves not only to express *existing* commitments,

> but also to provide the *framework* for the thinking by which we reason our way (in a manner appropriate to the specific subject matter) into the making of *new* commitments and the abandoning of old.
>
> (CDCM §103: 302–03)

There is a lot built into this thought, which comes near the end of a very long essay, so let's unpack it a bit. In order fully to understand Sellars' somewhat unorthodox approach to these matters, we need to understand how he thinks of inductive inquiry generally. He sums it up nicely in a catchword: "The motto of the age of science might well be: *Natural philosophers have hitherto sought to understand meanings; the task is to change them*" (CDCM §86: 288). Science is a rigorous method for systematically changing our language, and therefore our concepts, and therefore our abilities to cope with the world. In part, this means changing the inferential proprieties involving the terms and concepts we use. So take either a lawlike statement (perhaps, "To every action there is an equal and opposite reaction") or a statement containing an explicit causal modality (say, "If the books hit 451°F, necessarily, they'll combust"). Sellars tells us that, "scientific terms have, as part of their logic, a 'line of retreat' as well as a 'plan of advance'" (CDCM §86: 288).[14] These lines of retreat and advance are not themselves to be cashed out directly in terms of standing inferential proprieties, but rather in terms of susceptibility to subject-matter-dependent methodologies. There are particular ways to go about jeopardizing, extending, retracting, and revising the inferential proprieties that license appropriate use of empirical and theoretical terms.

Lawlike statements that play an explanatory role cannot be singled out by any syntactic markers I know of. Sellars points to the pragmatic status we accord them: we willingly project them into unobserved situations, precisely because we have come to assert/believe them on the basis (we hope) of appropriate, well-executed inductive methods. It is the background reliance on inductive methodology combined with sophisticated logico/mathematical analysis that distinguishes

> To every action there is an equal and opposite reaction

from

> To every thing there is a season, and a time to every purpose under heaven.

In understanding both these assertions, we grasp implicitly that there are different routes leading to acceptance of the claims and different routes involved in revising them or rejecting them. Similarly, the commitment made to each, to Newton's law and to the *Weltanschauung* of Ecclesiastes, has a different tenor to it. One expresses a law of nature; the other, perhaps, a natural law in a very different sense.

Inference from a generalization is rarely deeply explanatory on its own precisely because it is *thin* in contrast to inferences from lawlike statements that we take to have modal or subjunctive force. The latter are *thick* because the logical space in which they place things has extensive horizons and a significant infrastructure: a history and a future, appropriate and inappropriate methodologies, a penumbra of possibilities that, though not *described*, are *conveyed* in ways that enrich our understanding and guide our ability to go on. It is in the demand for explanation and the ability to meet that demand that the (regulative) *ideal* unity of a conceptual framework and the measure of its *real* success can be found.

B. Laws, Causation, Explanation, and Agency

The relation between the causal and the real I've assumed on Sellars' behalf turns out not to be so simple, for causation ultimately has a derivative status in Sellars' system. Again, our text is "Counterfactuals, Dispositions, and the Causal Modalities".

In Sellars' view, as I read him, causal notions properly so-called live in a very particular environment. It is an environment populated with thing-kinds, dispositions, capacities, what is *done,* and what *results*. This environment, of course, has its roots in the manifest image, but retains them even as the sciences grow, insofar as the sciences themselves use causal explanations. Yet Sellars' bottom line is that causal notions themselves will ultimately be dispensable. The laws of nature that science ultimately ends up with will have abandoned the categorial environment within which causation finds its niche. This would mean that the final scientific image of humanity in the world is not only beyond good and evil, but beyond causation as well. My argument, however, is that this simply goes too far. I've never fully understood Sellars' ideal process ontology (despite Johanna Seibt's best efforts), and I think more and more that there is good reason for that. A proper conception of ontology does not drive us to those extremes.

Causal explanations, in Sellars' view, essentially involve reference to the thing-kinds that the relevant objects belong to. Sellars has a rich notion of thing-kinds: terms for thing-kinds are not ordinary predicates but common nouns, and their grammar differs from adjectival and adverbial constructions. The role of kind terms in causal explanations is doubly complex, because things-kinds usually have causal properties among their distinguishing characteristics.[15] This is where we can pick up the thread of Sellars' thought, for he thinks it is important to distinguish "between the causal properties of a certain kind of thing, and the theoretical explanation of the fact that it has these causal properties" (CDCM §50: 263).

> For while causal generalizations about thing-kinds provide perfectly sound explanations, . . . it is no accident that philosophers have been tempted to think that such a phenomenon as salt dissolving in water

> must "at bottom" or "in principle" be a "lawfully evolving process" describable in purely episodic terms.
>
> (CDCM §50: 263)

But moving to such a process explanation that uses purely episodic terms means the abandonment of causation as we know and love it.

> Such an "ideal" description would no longer, in the ordinary sense, be in causal terms, nor the laws be causal laws; though philosophers have often muddied the waters by extending the application of the terms 'cause' and 'causal' in such wise that any law of nature (at least any nonstatistical law of nature) is a "causal" law.
>
> (CDCM §50: 263)

Any given mode of explanation might be such as to lead "*by its very nature*" to "new horizons", that is, to a new and different mode of explanation, "to new questions calling for new answers of a different kind" (CDCM §50: 263). The thing-kind-based generalizations that usually underwrite our explanations point beyond themselves, because thing-kinds *bunch* rather than *explain* causal properties. We must learn to "appreciate the promissory note dimension of thing-kind expressions". We begin to get beyond the framework of *manifest* causation by moving to the micro-level, at which many of the causal properties of molar things-kinds can be explained. But the narrative cannot stop there, Sellars believes, because "micro-theories themselves characteristically postulate micro-thing-kinds which have fundamentally the same logic as the molar thing kinds" (CDCM §51: 264). Sellars wants us, ultimately, to get out of the thing-kind business altogether.

I am not convinced that this is the way to go. For one thing, I do not have any clear conception of just what it is that Sellars thinks we're aiming at—and he rejects the effort to make good on his promissory note any time soon.

> The conception of the world as pure process, which is as old as Plato, and as new as Minkowski, remains a regulative ideal; not simply because we cannot hope to know the manifold content of the world in all its particularity, but because science has not yet achieved the very concepts in terms of which such a picture might be formulated.
>
> (CDCM §52: 264)

Just as Marx can only hint at the nature of the communist society to come, Sellars can only hint at the scientific millennium to come. But that is not actually a problem for me. What I find more problematic is his quick dismissal of his rivals. He tells us,

> Only those philosophies (New Realism, Neo-Thomism, Positivism, certain contemporary philosophies of common sense and ordinary usage, etc.)

> which suppose that the final story of "what there is" must be built (after submitting them to a process of epistemological smelting and refinement) from concepts pertaining to the perceptible features of the everyday world, and which mistake the methodological dependence of theoretical on observational discourse for an intrinsically second-class status with respect to the problems of ontology, can suppose the contrary.
>
> (CDCM §52: 264)

I think I have learned the lessons about the methodological (but not ontological) priority of the observation language, and I don't have any intentions of ensconcing some form of perceptual given in either my epistemology or my ontology. The considerations that his rivals hold on to but Sellars rejects are all on the input side of the epistemological process. We need to consider the output side as well. Sellars seems to have forgotten here his own belief that in the scientific image, we will not (and will not be able to) dispense with the language of individual and community intentions. I simply do not see how we can preserve that language in a framework of pure processes.

There are two problems I foresee here. The first, and perhaps lesser, problem is that it seems inevitable that a view of the world as a battery of pure processes governed by laws of nature will be computationally intractable in real time for creatures like us. The "bunching" that thing-kind concepts do turns out to be crucial: it introduces essential simplifications into our scheme that enable it to remain tractable while still offering "perfectly sound explanations." Empirically robust "bunches" of properties in thing-kinds reveal something significant about the world. Even if the "bunches" permit (or even demand) further explanation in terms of patterns of pure processes, we need not dismiss those patterns as not *really* real, much less as simply *unreal*.

If we take seriously the pragmatic criterion of ontological commitment I have recommended, then, however far science may progress, it will remain the case that we must count on and take account of those groups of higher-level patterns that form kinds, both natural and social. I am not falling into a permissive relativism here, for we must still respect the demand that such kinds as we recognize afford us "perfectly sound explanations" that hold up under scrutiny. The fact that they *do* afford such explanations tells us something significant about the ontological structure of our world.

Another way to put my point would be this: I, like Sellars, endorse a form of explanationism in ontology, but I do not endorse his explanatory *monism*. Good explanations come in many sizes and shapes, and each one tells us something important about the structure of the world.

The second, and to my mind, still more serious problem is that intentions have an ineluctable first-person reference to oneself as *agent* built right into them, and agents are a very special kind of thing. So the problem I see is that we cannot simply "abandon the framework of common sense and use

only the framework of theoretical science" (SRI: 189, in PP: 354), because we cannot *use* the framework of theoretical science, the framework within which some inferences and not others are legitimate, in which some descriptions and not others are proper, in which some procedures are productive and useful and others are wrong-headed or wasteful, unless we retain the conception of ourselves as epistemic agents. Were we fully to give ourselves over to a pure process view of reality, there would be laws of nature governing the occurrence of such pure processes, but there would be no *us* (as salient features of the environment) to recognize them. In the framework in which the language of individual and community intentions makes sense, persons are basic entities, fundamental unities; they could not remain such in a framework of absolute processes.

That agents and agency vanish in a pure process view of the world ramifies through Sellars' philosophy. Recall, if you will, Sellars' rather difficult discussion of the skeptical question in "More on Givenness and Explanatory Coherence," "Granted that we are in the framework [that produces our empirical knowledge], how can we justify accepting it?" Sellars tells us that the answer "lies in the necessary connection between being in the framework of epistemic evaluation and being agents" (MGEC ¶80: 190).

> 82. The answer is that since agency, to be effective, involves having reliable cognitive maps of ourselves and our environment, the concept of effective agency involves that of our IPM [introspective, perceptual, or memory] judgements being likely to be true, that is, to be correct mappings of ourselves and our circumstances.
>
> 83. Notice, then, that if the above argument is sound, it is reasonable to accept
>
> MJ5: IPM judgements are likely to be true,
>
> simply on the ground that unless they are likely to be true, the concept of effective agency has no application.
>
> (MGEC §§82–83: 190)

My argument is that in a pure process view of the world, the concept of effective agency makes no sense and has no application. But then claims to knowledge are ungrounded.

In my view, then, Sellars' belief that there is a regulative ideal of a system of pure processes that informs science ends up being self-stultifying; its achievement would undermine everything done to achieve it.

If my argument here is right, then we are at the doorstep of the next problem: showing that the kinds of thing-kinds that show up in our practical deliberations can also show up in good causal explanations, so that the justificatory structures of practical reason make contact with the explanatory structures of empirical and theoretical reason. But that's for another day.

Notes

1 W.A. deVries "Images, Descriptions, and Pictures: Personhood and The Clash" in *Sellars and His Legacy*, ed. James O'Shea (Oxford: Oxford University Press, 2016: 47–59. For references to Sellars' works I use the now industry-standard abbreviations, together with paragraph or section numbers (for print-version freedom) and page numbers to at least one specific edition. See the References section for full details.

2 There are, of course, complex issues here, but to the extent that we do not think that the biological category *homo sapiens* is either necessary or sufficient for personhood, we at least lean toward the idea that person is not a natural kind in the usual sense of that term. Is 'person' a term of some *social* science? Are the kinds recognized in the social sciences *natural* kinds? These are difficult issues about which there is much disagreement. Furthermore, to the extent to which the focus here is on Sellars, we have to admit that he did not say enough about the social sciences to be sure how he would answer these questions. But his claim that "the scientific image of man turns out to be that of a complex physical object" (PSIM ¶70, in SPR: 25; in ISR: 393) seems to imply that in the scientific image the vocabularies of the social sciences (and thus, the objects of their interest) are not basic and are eliminable in principle in favor of the more basic scientific vocabularies of the postulational sciences.

3 Robert B. Brandom, *From Empiricism to Expressivism: Brandom Reads Sellars* (Cambridge, MA: Harvard University Press, 2015), p. 195.

4 "Now it is clear that the above account of truth applies to all kinds of propositions, ranging from singular state-of-affairs intensions to the propositions of mathematics and even the propositions of practical discourse" (SM, Ch. IV, ¶30: 102).

5 "Since the term 'fact' is properly used as a synonym for 'truth' even in its most generic sense, so that we can speak of mathematical and even ethical facts . . ." (SM, Ch. V, ¶2: 116).

6 Brandom calls his modal realism "robust" (FEE: 195), but I would suspect that his teacher David Lewis would have called it pretty thin beer.

7 Brandom claims that "descriptive claims are subject to a distinctive kind of ought-to-be. . . . It ought to be the case that the content of a descriptive claiming [that contains a modal qualifier or subjunctive] stands in a special relation, which we might as well call "correspondence," to a modal fact, which it accordingly purports to state (and in case there is such a fact, succeeds in stating)" (FEE: 210). He calls this "semantic government of claimings by facts" (loc. cit.).

8 And indeed, Brandom recommends conceptual realism to us, which certainly would include categorial realism. See FEE: 204. But this is not clearly compatible with the idea that it is the *formal* structures of language that carry a particular ontological burden, for surely not all concepts are formal in the requisite sense, on pain of emptying the formal/material contrast of any interest or content.

9 Remember Sellars' late characterization of the Myth of the Given in FMPP, I: "To reject the Myth of the Given is to reject the idea that the categorial structure of the world—if it has a categorial structure—imposes itself on the mind as a seal imposes an image on melted wax" (FMPP, I "The Lever of Archimedes" ¶45: 12). Sellars does not doubt that language has a categorial structure, so the parenthetical remark that leaves open the question concerning the world must mean at least that there is no straightforward *a priori* inference from linguistic structure to worldly structure. And Sellars' belief that the manifest image is a functional, working image that can nonetheless be replaced by a different and better image with a different categorial structure also implies that there is no straightforward inference from linguistic or conceptual structure to worldly structure.

10 deVries, Willem A. "Naturalism, the Autonomy of Reason, and Pictures," *International Journal of Philosophical Studies*, 18: 3 (2010): 395–413.
11 In a new article, Rebecca Kukla makes what I take to be a similar point, though perhaps not identical: "To be real is just to be patterned, counterfactually robust, and resistant. Furthermore, nothing can be counterfactually robust or resistant except relative to some set of enacted strategies for coping and coordinating with it" ("Enacting Stances: Realism without Literalism": 6).
12 "Counterfactuals, Dispositions, and Causal Modalities" In *Minnesota Studies in the Philosophy of Science*, *Volume II: Concepts, Theories, and the Mind-Body Problem*, ed. Herbert Feigl, Michael Scriven, and Grover Maxwell (Minneapolis: University of Minnesota Press, 1958), pp. 225–308.] (hereafter CDCM), § 108: 306–07.
13 In case one thinks that it is their generality that is the problem, rather than their descriptiveness, note how Sellars continues: "Indeed, do we ever make descriptive 'all-' statements about the whole universe everywhere and every-when? As philosophers we can imagine ourselves doing so; but the idea that we are doing so every time we make an unrestricted lawlike statement is a product of bad philosophy" (CDCM §98: 299). This runs directly against Brandom's notion that e.g., the law of gravitation is descriptive.
14 Ultimately, this applies to *all* our terms, since no term is absolutely frozen in place with an unchangeable meaning, and we're learning all the time.
15 "Things belong to kinds which are characterized by clusters of powers, capacities, dispositions and propensities, or—to use a general term intended to cover all these, and more—causal properties" (FMPP, II: ¶2).

References and Further Reading

deVries, Willem A. "Naturalism, the Autonomy of Reason, and Pictures," *International Journal of Philosophical Studies* 18: 3 (2010): 395–413.
deVries, Willem A. "Images, Descriptions, and Pictures: Personhood and The Clash" in *Sellars and His Legacy*, James O'Shea (ed.) (Oxford: Oxford University Press, 2016), 47–59.
deVries, Willem and Triplett, Timm (KMG). *Knowledge, Mind, and the Given: Reading Sellars's "Empiricism and the Philosophy of Mind"* (Indianapolis, IN: Hackett, 2000), with the complete text of EPM.
Sellars, Wilfrid. (EPM) "Empiricism and the Philosophy of Mind," in *Minnesota Studies in the Philosophy of Science*, Vol. I, Herbert Feigl and Michael Scriven (eds.) (Minneapolis, MN: University of Minnesota Press, 1956), 253–329. Reprinted in SPR, ISR, and KMG.
Sellars, Wilfrid. (CDCM) "Counterfactuals, Dispositions, and the Causal Modalities," in *Minnesota Studies in the Philosophy of Science*, Vol. II, Herbert Feigl, Michael Scriven, and Grover Maxwell (eds.) (Minneapolis, MN: University of Minnesota Press, 1957), 225–308.
Sellars, Wilfrid. (PSIM) "Philosophy and the Scientific Image of Man," in *Frontiers of Science and Philosophy*, Robert Colodny (ed.) (Pittsburgh, PA: University of Pittsburgh Press, 1962), 35–78. Reprinted in SPR and ISR.
Sellars, Wilfrid. (SPR) *Science, Perception and Reality* (London: Routledge and Kegan Paul, 1963; reissued by Atascadero, CA: Ridgeview, 1991).
Sellars, Wilfrid. (SRI) "Scientific Realism or Irenic Instrumentalism: A Critique of Nagel and Feyerabend on Theoretical Explanation," in *Boston Studies in the Philosophy of Science*, Vol. II, Robert Cohen and Max Wartofsky (eds.) (New York: Humanities Press, 1965), 171–204.

Sellars, Wilfrid. (SM) *Science and Metaphysics: Variations on Kantian Themes*, The John Locke Lectures for 1965–66 (London: Routledge & Kegan Paul, 1968; reissued by Atascadero, CA: Ridgeview, 1992).
Sellars, Wilfrid. (MGEC) "More on Givenness and Explanatory Coherence," in *Justification and Knowledge*, George Pappas (ed.) (Dordrecht, Holland: D. Reidel, 1979), 169–182.
Sellars, Wilfrid. (FMPP) "Foundations for a Metaphysics of Pure Process" (The Carus Lectures) *The Monist* 64 (1981): 3–90.
Sellars, Wilfrid. (ISR) *In the Space of Reasons: Selected Essays of Wilfrid Sellars*, Kevin Scharp and Robert Brandom (eds.) (Cambridge, MA: Harvard University Press, 2007).

10 Natural Truth

Danielle Macbeth

Impressed with how differently people from different cultures live and even think, the ancient Greek Sophists argued that what a given person takes to be true is really nothing more than a convention of his or her community. And since there is no independent, neutral vantage point from which to adjudicate such contradictory claims, it must be concluded that all truth is relative, that man is the measure of all things—just as Protagoras said. Neither Socrates, nor Plato, nor Aristotle found the argument compelling, and in their efforts to achieve knowledge of what is was born a project that within a few thousand years would grow into that of the exact sciences. The finest fruits of these sciences are the same not just for human beings but for *all* rational beings.

Although not in our everyday sensory perception of things, we can through the progress of the exact sciences achieve what we can think of as absolute truth, truth that is the same for all rational beings. Sellars (PSIM, 407) appears to have concluded from this (promised) achievement that our community, having begun parochially, as my or your tribe, and progressed "almost," he thinks, to "the 'brotherhood' of man," is poised to become the community of all rational beings, regardless of the kinds of bodies they happen to have. But although we and (say) the rational Martians may converge in our mathematical physics because and insofar as that science has been purged of all the contingencies of our bodies, histories, and cultures, it hardly follows that we and the rational Martians could live together in a single community. Our forms of life may be just too different to sustain meaningful discourse outside of highly abstract mathematics—or even at all.[1] But what should we say of those with whom we *do* share a biological form of life, though perhaps not a linguistic and cultural one? Given the role that, as Sellars argues, acculturation plays in the shaping of our everyday lives, it is natural to think with the Sophists that any notion of truth appropriate to our everyday lives—as contrasted with findings in the exact sciences—must be relativized to one's linguistic and cultural life form.

We know that one's acculturation profoundly affects the way one perceives and thinks about reality. For example, whereas basic laws of logic such as the law of non-contradiction were discovered very early in the

history of the West and continue to provide an essential guide to what Westerners judge to be correct thinking, it is dialectical reasoning by which to reconcile, or transcend, or even accept contradiction that dominates in the East. If one is Chinese or Japanese, one will tend to seek to explain things by appeal to relationships and context. American and European thought is more analytical, focused on individual objects and their properties, and on the laws that are taken to govern them. If, for example, one shows a clip of some swimming fish and other underwater objects in a pond to, say, some Japanese and American students, the American students will report first on the fish, the Japanese on the pond. And the Japanese students will be much more attentive to relationships among the various objects depicted than the Americans. The two groups of students *see* things in radically different ways. Sellarsians will not be surprised to learn that these deep cognitive differences are widely held to be "due primarily to differences in language and writing systems" (Nisbett et al. 2001, 304). Cultural relativism seems inevitable.

The idea of *natural truth*, modeled on the idea of natural goodness in ethics, is the idea that some truths, while not the same for (that is, available to be grasped by) *all* rational beings—as the truths of mathematics and fundamental physics are, at least in principle—are nevertheless valid for (available to) all *human* beings, all rational beings with our sort of body and form of sensibility. Natural truth is incompatible with cultural relativism. As things stand, of course, people from radically different cultures do perceive and think in very different ways about the perceptible world we live in. The question is: are they *right* to do so, or is there in fact a way that the humanly perceptible world *is*, however, we, or they, take it to be, a way we—rational animals that we are—*ought* to take it? Are there natural truths?

The question of natural truth is the question whether the perception of us rational animals is properly speaking a power of knowing. It may seem obvious that it is not on grounds that our perceptual experiences are caused in us by brain activity that is itself caused, at least in ordinary cases, by the impacts of bits of reality on our sense organs. This objection is, I think, founded on a radically mistaken conception of the nature of perception. As we will begin to see here (and as I argue at length in my 2014 book *Realizing Reason*), we must understand perception not as an appearance caused in us but instead as a mode of access to things in the environment. Much as the emergence of living beings, plants, say, is at the same time the emergence of significances for various natural stuffs—water, for instance, is realized as nourishing—so with the emergence of animals capable of perceiving are realized various sensory properties of things. Perception is a mode of cognitive access to the world around us; the fact that it is sensory does not show that it is not a power of knowing, that there are no natural truths.

What *does* seem to show that perception is not a power of knowing, that there are no natural truths, is what appears to be the *fact* of cultural relativity. It is *unquestionable* that our perceptual experiences of things are

profoundly shaped and colored by our acculturation into one or other of a great variety of natural languages. Thus it can seem simply inconceivable that there might be some one (natural) truth about the world as it is revealed in our sensory perception. Nevertheless, my aim is to show that there *are* natural truths, and in particular to show that Sellars provides us all the resources we need to *establish* that there are natural truths. We will consider, first, two themes from Sellars that together can seem to suggest that our perceptual capacities do *not* provide us a means of knowing of how things are. When, however, we add two more Sellarsian themes to the mix a very different picture emerges, one according to which, as I will argue, there is and must be natural truth.

The first Sellarsian theme is the very familiar one we have already encountered: that rationality is not a biological phenomenon but instead an essentially social and cultural one. The human animal comes to be rational only through being initiated into a sufficiently developed linguistic culture. To suppose that we could be born into the space of reasons or could become rational in the course of our merely biological maturation would be to fall into what Sellars unmasks as the myth of the given.

As a living organism is an instance of a form of life and cannot be adequately understood except as such an instance, so a rational animal is an instance of a rational form of life and cannot be adequately understood except as such. In both cases the form of life provides a kind of model and standard. Consider, first, the more general case, that of a living organism. In order to understand, indeed, so much as to *identify* what it is doing, or what its parts are, or are for, one needs to know the *kind* of thing it is, and how things characteristically are for such a kind of thing. It is only in light of such knowledge about the form of life that one can judge of *this* instance that it has, say, wings though unfortunately they are so deformed in this case that the thing cannot fly as it should. So it is with a rational animal. We need to identify the language that it speaks, and more generally, the rational life form it is, and to know as well how things characteristically are for such a life form, for example, what various words and phrases mean in the language that this life form speaks. And in this case as well, it is only in light of such judgments about the form of life that one can judge of particular instances, say, of the sounds that a particular person now utters, what words they are and what the person is thereby saying. As Sellars (PSIM, 384–85) puts the point, "there is no thinking apart from common standards of correctness and relevance, which relate what *I do* think to what *anyone ought to* think. The contrast between '*I*' and 'anyone' is essential to rational thought." "It is the linguistic community as a self-perpetuating whole which is the minimum unit in terms of which conceptual activity can be understood" (LTC, 64).

The second theme of concern to us here, defended first in "Inference and Meaning" and then again in "Is There a Synthetic 'A Priori'?", is that there is nothing to conceptual meaning that is not determined by material and formal rules of inference: "the conceptual meaning of a descriptive term is

constituted by what can be inferred from it in accordance with the logical and extra-logical rules of inference of the language (conceptual frame) to which it belongs" (ITSA, 317; cf. IM, 25–26). Although, according to Sellars, language *use* (and hence, meaning in a broad sense) involves both language entry transitions and language exit transitions in addition to language-language moves, that is, inferences, it is *only* the inferential moves that contribute to what Sellars refers to as conceptual meaning. Sellars' grounds for this thesis are clear: although there is an obvious sense in which a rule of inference, whether material or formal, is a *rule*—something that sets out what one ought or is permitted to do in certain specified circumstances—language entry transitions (and language exit transitions) cannot in the same sense be governed by rules.

> [To characterize] the learning to use a language or system of concepts as learning to use symbols in accordance with two types of rules: (a) rules of syntax, relating symbols to other symbols; (b) semantical rules, whereby basic factual terms acquire extra-linguistic meaning . . . involves a radical mistake. A rule is always a rule for *doing* something in some circumstance. Obeying a rule entails recognizing that a circumstance is one to which the rule applies. If there *were* such a thing as a semantical rule by the adoption of which a descriptive term acquires meaning, it would presumably be of the form 'Red objects are to be designated by the word "red".' But to recognize the circumstances in which this rule has application, one must already have the concept of red!
>
> (ITSA, 312; cf. IM, 25)

Despite the fact that, as Sellars recognizes, an English speaker ought to be disposed to utter the word 'red' in the presence of red things, Sellars denies that a word such as 'red' has empirical meaning in virtue of "its role as a conditioned response to red things" (1958b, 314). There is, we will later see, a kind of rule associated with one's disposition to utter, say, 'red' in the presence of red things, but it is not, according to Sellars, a rule in the same sense as that in which an inference rule, whether material or formal, is a rule. It is just this that Sellars emphasizes in saying that there is nothing to the conceptual meaning of a word that is not given by the formal and material rules of inference in which it figures. That English speakers ought to be disposed to utter the word 'red' in the presences of red things simply does not have the form of a *rule* (to do something in certain circumstances). As the point is put in "Some Reflections on Language Games" (§36), whereas intralinguistic moves are *inferences* according to syntactical rules, whether material or formal, language entries are instead *observations*.[2]

Standard logic does not recognize material rules of inference. Sellars holds that language is unintelligible without them: "Material rules are as *essential to meaning* (and hence to language and thought) as formal rules,

contributing the architectural detail of its structure within the flying buttresses of logical form" (IM, 7). As Sellars puts the point some years later:

> It is only because the expressions in terms of which we describe objects, even such basic expressions as words for the perceptible characteristics of molar objects, locate those objects in a space of implications, that they *describe* at all, rather than merely *label*. The descriptive and the explanatory resources of language advance hand in hand; and to abandon the search for explanation is to abandon the attempt to improve language period.
>
> (CDCM, 306–07)

To *describe* rather than merely label requires just those capacities that enable explanation, that is, rules of material inference that track, or purport to track, causal and other (necessary) relations. And it does so because in describing one is answerable to how things are however one takes them to be. It is intelligible that one *is* describing, then, only if one is *responsive* to how things are, answerable to the norm of truth. It must be in one's power to get it right, and hence in one's power to correct one's errors. And this requires in turn that there be material as well as formal rules of inference because only so can there be conflict and second thoughts.

The basic idea is very simple, and very familiar. Suppose that I see (or at least take myself to see) a thing to be at once F and G despite the fact that there is a material rule in my language to the effect that being F entails being not-G. That is, I can infer, given that the thing is F, that it is not G, even though it appears to be G. Although I find myself wanting to *report* observationally that the thing is G, I am also inclined to state, as the conclusion of an inference, that it is not G—either that or that it is not F, or that after all it cannot be inferred from something's being F that it is not G. In short, I have an inconsistent triad: the thing is F, it is G, and being F entails being not-G. Something has to give (and usually, in actual practice, it is obvious what). This is what material rules of inference give us, not only the capacity to explain but also the capacity for second, better thoughts both about how things are and about what is a reason for what. And it is this capacity for second thoughts, for self-correction, that ensures that we are describing, and on occasion misdescribing, rather than merely labeling, that in our talk we are answerable to things as they are.

A language, according to Sellars, is a "system of formal and material rules of inference." And "there are," he holds,

> an indefinite number of . . . [such] systems . . . each one of which can be regarded as a candidate for adoption by the animal which recognizes rules, and no one of which has an intuitable hallmark of royalty. They must compete in the market place of practice for employment by language users, and be content to be adopted haltingly and schematically.
>
> (IM, 26)

It is clear that Sellars himself thinks that it is some future language of science that ultimately will be the language we need in order to describe and explain things as they actually are. But although Sellars recognizes "an indefinite number" of languages, he recognizes only two "images," two "ways of experiencing the world" (PSIM, 373). It is, for example, in the manifest image, in that way of experiencing, that we perceive a lawful relation among the pressure, temperature, and volume of a gas. This observed regularity is then explained in the scientific image by the kinetic theory of gases, which postulates unobservable little atoms and molecules flying about and bumping into one another and the walls of the container of gas. The behavior of these postulated unobservables in the scientific image explains the observed regularity in the manifest image.

It is furthermore clear that Sellars thinks that only one of the two images can be right. One can follow the lead of the philosophers of 'common sense' and 'ordinary usage' who accept "the manifest image as *real* . . . [and] attempt to understand the achievements of theoretical science in terms of this framework, subordinating the categories of theoretical science to its categories" (PSIM, 387). Or one can follow Sellars' lead (EPM §42), holding instead that "in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not." Although, Sellars (PSIM, 382) thinks, "there is truth and error with respect to" the manifest image, because it "has a being which transcends the individual thinker," in the end that way of experiencing is to be rejected as false.

According to Sellars, it is our scientific theories and rationally constructed scientific languages that are the medium within which to achieve an adequate cognitive relation to reality, not our natural languages. Given that a language is a system of (formal and material) rules of inference, and the medium of our cognitive commerce with reality, we can see the project of science as that of discovering the system of formal and material rules of inference that will enable us to see (better: *think*) reality as it is in itself. Although we *begin* with our various colloquial languages, ultimately we will achieve the language that is the same for all rational beings insofar as they are rational, whatever their biological and cultural forms of life. There are, on this Sellarsian view, no natural truths to contrast with the absolute truths of a completed science.

Language use is clearly a rule-governed activity; there are norms, standards of correct word usage. But according to Sellars, not all rules of language are rules of action, rules enjoining one to perform a certain action in particular circumstances, what Sellars calls ought-to-dos. We have seen that inference rules are such rules, but the rules governing what Sellars calls language entry transitions and language exit transitions are not because, as we already saw for the case of language entry transitions, to recognize the circumstances as the circumstances in which one ought to describe a thing as, say, red, is already to see the thing as red, to have tacitly so taken it.

But seeing something as red is also not merely a matter of being caused to act in a certain way. Seeing something as red is normative: one ought to see red things as red, and generally ought to see things as they are. Hence we need to recognize, in addition to rules of action, ought-to-dos, also what Sellars calls rules of criticism, ought-to-bes. There is and can be no rule of *action* according to which "(*ceteris paribus*) one ought to say 'this is red' in the presence of red objects in sunlight"; but there *is* a rule of *criticism* according to which "(*ceteris paribus*) one ought to respond to red objects in sunlight by uttering or being disposed to utter 'this is red'" (1969, 63). One ought to be that sort of person, the sort of person who is disposed in this way, which implies in turn an ought-to-do: people—others when one is young but also oneself when one has become a competent speaker—ought to bring it about that one *is* that sort of person. But if so then there is a sense in which, although perception is a receptive faculty, something with respect to which one is passive, one is nonetheless responsible for seeing things as they are—not *directly* responsible, but responsible nonetheless. This is our third and penultimate Sellarsian theme, that to understand language and thought we need to recognize not only rules of action governing inferences but also rules of criticism governing one's perceptual experience of things.

My final Sellarsian theme is one that gets little attention, though it is, I think, absolutely fundamental. It is that inquiry "is rational, not because it has a *foundation* but because it is a self-correcting enterprise which can put *any* claim in jeopardy, though not *all* at once" (1956 §38). We know that the myth of the given is the idea that something non-conceptual, something outside the space of reasons, might nonetheless serve as a reason. What this last Sellarsian theme tells us is that there is another pitfall to be avoided as well. For it is all too easy for one, having once learned that the given is a myth, to recoil to a position that is equally mythic, what we can think of as the myth of the taken, the idea that we are not after all answerable to how things *are* but only to the tribunal of experience, that is, the appearances of things to creatures like us, which, as long as we recognize the role of concepts in it, does not land us in the myth of the given. If all we are answerable to is the tribunal of experience, if we are not answerable to things as *they* are, however we experience them to be, then we are not answerable in any way that counts as rational. It must be *things as they actually are* that bear rationally on our empirical thinking if what we are doing is to be recognizable as thinking. It follows directly that the idea that there are things in the space of reasons that are unrevisable in principle (hence not answerable to things as they are—as mere appearances are not so answerable) is just as much a myth as the given is. Indeed, this second myth seems in fact to constitute yet another episode in the myth of the given insofar as in this myth one aims to hold onto the idea that everything in the space of reasons is conceptually articulated while denying what seems to be a consequence of that idea, that anything in the space of reasons can be made the target

of one's critically reflective scrutiny, and corrected. Suppose that there were something we thought that was unrevisable in principle, something we just *had* to accept as true. The fact that we were so inclined to think would be no reason at all to think that what we think is actually true. If, on the other hand, *anything* we think can be called into question as reason sees fit, can be subjected to critically reflective scrutiny, then its *withstanding* such scrutiny *is* reason to take it to be true, at least until we find new reasons to cast it into doubt.

On Sellars' account, we have seen, the whole way of experiencing, the *framework* that he calls the manifest image, is to be rejected as false. And we are clearly supposed to take this as a move in the space of reasons. That framework was subjected to critically reflective scrutiny and is to be jettisoned in favor of the scientific image. It is, however, one thing to reject some claim within a framework, something else entirely to reject a whole framework. To reject the *framework* of the manifest image is to say, in effect, that everyday experience is a mere appearance caused in us, that it is not really real. To reject the framework of the manifest image is to claim that the observed regularities, say, among the pressure, temperature, and volume of a gas are not actually *true*, not in fact answerable to how things *are*, but only an appearance caused in us, an appearance to be explained by science. And *that* is just the myth of the taken, the myth that in the end our thinking is answerable only to things as we take them to be, as they show up in our experience—however they in fact are.

In Sellars' scientistic picture according to which our everyday claims do not actually describe how things are, everyday experience is a mere appearance caused in us, an appearance that is ultimately to be explained by our best scientific theories. But if everyday experience *is* such an appearance, if it is not itself answerable to how things actually are, then it is Sellars who falls into the myth of the taken. If all our scientific theories answer to is our experience of things conceived as a mere appearance, if those appearances are not *themselves* answerable to things as they are, then the practice of science is not and cannot be a *rational* endeavor. It is *only* if our ordinary sensory perceptions of things are themselves rational—that is, answerable to the norm of truth—that we can make intelligible the project of science as that of determining how things are in an absolute sense the same for all rational beings. Everyday experience is not a mere appearance to be explained by science; it is our first, sensory mode of cognitive access to reality, one that can be supplemented, though never supplanted, by the subsequent, non-sensory, purely rational mode of cognitive access that we achieve in the exact sciences. Everyday sensory experience is the domain of natural truth. And in this domain of truth as in any other, although where we *start* is completely contingent and arbitrary, a function of accidents of our history, context, biology, and environment, how we go on need not be. And it will not be just insofar as we exercise our powers of reason, insofar as we are critically reflective, and self-correcting.

We have seen that not only is nothing given for thought from outside the space of reasons, *nothing* is unrevisable in principle. *Anything* that we think can be called into question and corrected. Of course, at any particular moment in history some errors and confusions will be less accessible to rationally reflective criticism than others. Still errors *are* errors and ultimately ought to be corrected. And this is intelligible precisely because language is socially rather than biologically evolved; only if it is completely contingent, a matter of historical accident, how one takes things to be—at least at first—can one correct those errors, come to see things as *they* are. Even our everyday understanding of things can be improved, and it can be improved because, first, there are material rules of inference that enable contradictions to arise in what I seem to see to be so and am apparently entitled to infer is so, and second, there are rules of criticism that entail in turn rules of action, the upshot of which is that I am responsible not only for what I do, for instance, infer, *but also for what I perceive*. I am a rational being answerable to the norm of truth. I am responsible *for* all that I think and *to* things as *they* are. And from this natural truth follows.

Consider, for example, the fact that we have, all of us, been acculturated to give greater credence to an utterance by a male than to the exact same words, uttered with the same degree of assurance and air of authority, spoken by a female. The male just *sounds* smarter, more interesting and creative, more knowledgeable, and more authoritative. Similarly, black males in the United States are all too often perceived by whites in the United States as particularly dangerous, threatening, and violent, even in cases in which there is no discernable difference between the appearance and behavior of the black male and the appearance and behavior of his white counterpart—save, of course, for the fact that the one is black and the other white. And these *are perceptions*, as manifest to a perceiver as the blue of the sky or the five fingers on my hand. One does not in such a case *infer*, even unconsciously, that things are thus and so; one literally sees, or at least seems to see, them to be so. And yet, there seems in these cases *clearly* to be a fact of the matter that we are getting wrong. The fact of being female or black does not provide sufficient grounds for the perceptions one has in these cases. And, if Sellars is right, the fact of the matter in such cases is nevertheless one to which we are answerable. One ought to see things, in this instance, other people, *as they are* and not merely as one has been socialized, acculturated to see them. And one ought to do this because although perception is passive, one can—and insofar as one is rational ought to—bring it about that one is the sort of perceiver who sees things correctly, as *they* are. Now we need to generalize the point.

As human beings we share a certain biological form of life; we are, or at least begin as, *animals* of one certain sort. And our languages, we have seen, are utterly contingent; anything we think could in principle be revised. We also know that we could, each of us, learn to see things as those in other cultures do by immersing ourselves in their languages and cultures. I could,

for example, come to be able to see the depiction of the fish in the pond not only in terms of objects with their properties as I as a Westerner am prone to, but also and primarily in terms of context and relationships as an Easterner does. Now if I am merely looking at a picture, a depiction, then there would seem to be nothing to choose between the two ways of seeing; I can look at the picture the one way and I can look at it the other. Neither way of regarding the picture, by contrast with the other, is *true* or *correct*. But the same cannot be said when we turn to the things themselves. If I am looking not at a depiction of fish in a pond but at some *actual* fish in a pond then there *is* some fact of the matter how they ought to be regarded, whether atomistically, or relationally and contextually, or in some other way. *Can* one adequately perceive fish, perceive them *as they are*, if one considers them independently of their environment? The question seems obviously to be a good one, a question about the fish, however I have been acculturated to see them. But if that question is a good one then there is natural truth.

Perhaps it will be objected, again, that in all such cases it is merely a matter of taste, or of convention, how we regard this or that thing. I do not see how it can be. Suppose that you like chocolate ice cream best and I prefer the taste of vanilla ice cream. I can in that case *understand* your liking chocolate ice cream better than vanilla ice cream, but I cannot *experience* what you experience—namely, chocolate ice cream tasting better than vanilla ice cream. If I could do that then I *would* like the taste of chocolate ice cream better. There is no natural truth about which of the two tastes best; it is only a matter of taste. But again, one and the same person *can* learn to see the fish now atomistically and now holistically. And yet the two views would seem to be incompatible insofar as the former view is non-relational and the latter essentially relational. They cannot both be right—or so a student of Aristotle might argue. A student of Nagarjuna might suggest instead that both perspectives are needed fully to account for how things are with the fish. The point remains that we cannot merely leave each to his or her view of things. The question as to the nature of fish is not a question about how things seem to one, as questions of taste are. Nor, obviously, is it merely a matter of convention how things are with the fish, as it is a convention which side of the road one should drive on. The question about the fish is a question about how things are however we perceive them to be, how rational animals such as us ought rationally to perceive and think about fish. It is a question of natural truth.

We saw that Sellars envisages a community of all rational beings. There can be no such community. Rational beings with a radically different form of life from our own would be unintelligible to us—as we would be to them. But if, as I have argued, there is natural truth then there *can* be a community of human beings, what Sellars describes as "the 'brotherhood' of man." Sellars, we saw, thought we had nearly achieved such a brotherhood. No one would think so today. The cultural differences that divide us seem to be as deep, and as divisive, as they could possibly be. All the more reason, then, to

take up the task of determining how we *ought* to think about the everyday, perceptible world. We are, of course, not starting from scratch, any more than we ever are. We begin where we are, with all that we think we know and understand, and all that is available to be learned. What we need is the sort of rational, reflective criticism that is the hallmark of reason. We need to *think* about the different ways people have come to understand reality, and to come to better second thoughts about how things actually are. We need to attend to natural truth.[3]

Notes

1 And if we could have no meaningful discourse with them then perhaps we could not recognize them as rational at all. They might be rational nonetheless. See my (2014, Ch. 1).

2 Although I cannot pursue the point here, this distinction that Sellars draws between what he refers to as the conceptual meaning of a term, constituted by rules of inference, and observations, has very suggestive affinities with Frege's distinction between the sense (*Sinn*) expressed by a term and that term's signification (*Bedeutung*).

3 My thanks to David Pereplyotchik for helpful suggestions for improvements.

References and Further Reading

Macbeth, Danielle 2014 *Realizing Reason: A Narrative of Truth and Knowing*. Oxford: Oxford University Press.

Nisbett, R. E., K. Peng, U. Choi, and A. Norenzayan 2001 Culture and Systems of Thought: Holistic Versus Analytic Cognition. *Psychological Review* 108 (2): 291–310.

Sellars, Wilfrid 1953 Inference and Meaning. Reprinted in *In the Space of Reasons*.

——— 1954 Some Reflections on Language Games. Reprinted in *In the Space of Reasons*.

——— 1956 Empiricism and the Philosophy of Mind. Reprinted in *Science, Perception and Reality*.

——— 1958a Counterfactuals, Dispositions, and the Causal Modalities. In *Minnesota Studies in the Philosophy of Science*. Vol. II, ed. H. Feigl, M. Scriven, and G. Maxwell. Minneapolis: University of Minnesota Press: 225–307.

——— 1958b Is There a Synthetic 'A Priori'? Reprinted in *Science, Perception and Reality*: 298–320.

——— 1960 Philosophy and the Scientific Image of Man. Reprinted in *In the Space of Reasons*: 369–408.

——— 1963 *Science, Perception and Reality*. London: Routledge and Kegan Paul.

——— 1969 Language as Thought and Communication. Reprinted in *In the Space of Reasons*: 57–80.

——— 2007 *In the Space of Reasons: Selected Essays of Wilfrid Sellars*, ed. K. Scharp and R. B. Brandom. Cambridge, MA: Harvard University Press.

11 Does Brandom's "Kant-Sellars Thesis about Modality" Undermine Sellars' Scientific Naturalism?

Dionysis Christias

1. Introduction

In his new book *From Empiricism to Expressivism: Brandom Reads Sellars* (2015), Brandom offers a new argument against the viability of Sellars' scientific naturalism, as the latter is famously expressed in the *scientia mensura* principle according to which "in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is and of what it is not that it is not" (Sellars 1997, §42). A very interesting feature of this critique is that it utilizes explicitly and exclusively *Sellarsian* premises in order to reach its conclusion, suggesting thereby the possibility of a serious internal tension within Sellarsian philosophy itself. More specifically, Brandom uses what he calls the "Kant-Sellars thesis about modality" in order to argue against Sellars' scientific naturalism, which privileges the descriptive and explanatory resources of the scientific image over those of the manifest image in matters ontological. Brandom shows that the Kant-Sellars thesis about modality implies an analogous thesis about identity, namely the "Kant-Sellars thesis about identity." And this latter thesis, implying as it does that manifest-image objects cannot be identical to scientific-image objects, violates the *scientia mensura* principle, since, according to Brandom, this principle amounts to the claim that manifest-image objects exist only if they are construed as being identical to objects specifiable in the language of eventual natural science. I shall argue that, while it is true that the Kant-Sellars theses about modality and identity do violate the *scientia mensura* principle, *as Brandom understands it* (i.e., as an identity relation between the objects of the manifest and the scientific image), they are in fact fully consistent with the latter as *Sellars* understands it. This is because the relation between manifest-image objects and scientific-image objects is explanatory, and an explanatory relation is not one of identity. The fact that manifest-image descriptions are ultimately to be explained in scientific-image terms implies that phenomena described in manifest-image terms do not really exist, in the sense that they can be considered as implicated in causally efficacious phenomena (i.e., as providing adequate explanations) only if they are *recategorized* and ultimately understood in *different*, *scientific*-image terms.

2. The Kant-Sellars Thesis about Modality

But what exactly is the Kant-Sellars thesis about modality, and how can it be used to undermine Sellars' scientific naturalism? This thesis is based on Sellars' contention to the effect that offering any empirical description, attributing any empirical property, involves commitments as to what *would* happen to what is so described under various circumstances (i.e., what would be true of it *if* various other claims *were* true). Sellars argues that what distinguishes description from mere labelling is precisely that circumstances of appropriate application, sufficient only for labelling something,[1] are paired with *consequences* of such application. Thus describing something places it in a space of implications, which inferentially articulates the content of the description. As Sellars himself puts this point:

> Although describing and explaining (predicting, retrodicting, understanding) are *distinguishable*, they are also, in an important sense, *inseparable*. It is only because the expressions in terms of which we describe objects, even such basic expressions as words for the perceptible characteristics of molar objects, locate these objects in a space of implications that they describe at all, rather than merely label. The descriptive and the explanatory resources of language advance hand in hand.
>
> (Sellars 1957, §108)

A necessary condition for an expression having a descriptive role in language is being situated in a space of implications, which, moreover, for Sellars, must be counterfactually robust ones. That is, they must *remain* good under various merely hypothetical circumstances; otherwise the putatively "descriptive" term could not be consistently applied to new cases. And it is an essential feature of this space of (counterfactually robust) implications that endorsing and inference in it is something that can be appealed to in *explaining* the applicability of one description by the appealing to the applicability of another (e.g., Brandom 2015, 182). Notice moreover that the inferences in this space of implications always include inferences that involve collateral premises or auxiliary hypotheses not drawn exclusively from one's current commitments.[2] (What Sellars means by 'explanation' is understanding the applicability of some descriptions as inferable from the applicability of others according to just this kind of-subjunctively robust and essentially nonmonotonic inference.) *Every* empirical descriptive concept has modal consequences. That is, its correct application has necessary conditions that would be expressed explicitly using subjunctive conditionals, and hence depends on what is true in other possible worlds besides the one in which it is being applied. For example, consider the following ordinary descriptive sentences:

1) The chunk of iron has a mass of 1 kilogram.
2) That lion is sleeping lightly.
3) The paint patch is red.

(1) cannot be true unless

(1') "A force of 1 Newton *would* accelerate the chunk of iron at 1 meter/second^2" is also true.

(2) has as necessary conditions that

(2') some moderate stimulus (e.g. a sufficiently loud noise, bright light) would wake the lion.

And (3) entails

(3') "the patch would look red under standard conditions, and would look brown to a standard observer under green light."

Thus, an important consequence of the Kant-Sellars thesis about modality is that one who understood none of the subjunctive implications one was committing oneself to by applying the descriptive terms 'mass', 'lion', or 'red' could not count as grasping the concepts they express.

3. The Kant-Sellars Thesis about Identity as an Argument against the *Scientia Mensura* Principle

Now, in what way is the Kant-Sellars thesis about modality related to Sellars' scientific naturalism, and how can it undermine the latter? Schematically put, Brandom takes it—and in this I think he is right—that the Kant-Sellars thesis about modality has interesting consequences about *identity* claims. Moreover, Brandom also construes the relation between the manifest and the scientific image, as described in the *scientia mensura* principle, as an *identity* relation between objects of an observational-correlational framework, which ultimately present mere appearances, and objects of a theoretical-postulational framework, which are about "things-in-themselves," "what really exists." That is to say, Brandom interprets the *scientia mensura* principle as entailing that descriptive terms from the manifest image refer to things specifiable in descriptive terms from the scientific image, if they refer at all; if some nonscientific descriptive term refers to anything real (rather than presenting a mere appearance), it is only because it corefers with some scientific descriptive term. But coreference of terms is *identity of objects*. In this way, Brandom construes the claim of the *scientia mensura* principle as follows: to exist requires being identical to some object specifiable in the language of eventual natural science (Brandom 2015, 62). Brandom goes on to argue that the *scientia mensura* principle is in conflict with the Kant-Sellars thesis about identity. But before challenging this latter Brandomian claim, let us first briefly explicate the Kant-Sellars thesis about *identity* and its connection with the Kant-Sellars thesis about modality.

According to the Kant-Sellars thesis, the correct application of an empirical descriptive concept has necessary conditions that would be expressed explicitly using subjunctive conditionals, and hence depends on what is true in other possible worlds besides the one in which it is being applied. But this thesis is incompatible with extensionalism about identity (i.e., the well-known thesis, famously propounded by Quine, according to which identicals are indiscernible only with respect to 'non-modal' properties). This is because the defining feature of extensional predicates/properties is that what they apply to in a given possible world, for instance, the actual world, depends only on what is true at that world. They are in this sense *modally insulated*, in that their conditions of applicability (what they describe) are insulated from modal facts (i.e., facts about what *would* happen if some other circumstances *were* different). Now, as is evident, this extensionalist view is in direct conflict with the Kant-Sellars thesis about modality, according to which every empirical descriptive concept has modal consequences. Thus, *pace* extensionalism about identity, if one accepts the Kant-Sellars thesis about modality, one cannot consistently restrict the properties with respect to which identicals must be indiscernible to descriptive properties that are 'nonmodal', on pain of emptying those descriptive properties of any determinate content. As Brandom himself puts this anti-extensionalist point, on behalf of Sellars:

> All descriptive predicates have subjunctively robust consequences because . . . being located in a space of such explanation-supporting implications is just what distinguishes *descriptions* from mere labels. Describing something in the actual situation *always* involves substantial commitments as to how it *would* behave, or what else *would* be true of it, in other *possible* situations.
>
> (Brandom 2015, 67–68)

As we saw above, this is so even in the case of descriptive terms that seem to be the best candidates for modally insulated predicates or properties, such as 'mass', 'lion', 'red', etc. Thus, from this Kantian-Sellarsian point of view, extensionalism is seen as a view that ends up draining basic descriptive terms of any *determinate* content, undermining thereby the extensionalist construal of claims about identity.

To sum up, one consequence of the Kant-Sellars thesis about modality (i.e., that there are no modally insulated properties) concerns claims about identity: the properties with respect to which identicals must be taken to be indiscernible cannot be restricted to modally insulated ones. But as Brandom rightly notices, allowing modal properties into the class of properties with respect to which identicals must be indiscernible has radical consequences. If, as the Kant-Sellars thesis about identity has it, identity requires indistinguishability by modal-dispositional properties, then it follows that two items that differ in their modal properties cannot be identical.

More importantly, this is so with every item that is identified and individuated on the basis of the *functional role* it plays within a system of other such items. Consider, for example, the relation between a valve and the particular piece of metal machinery from which it is constructed or the relation between a statue and the lump of clay or marble from which it is made. Although the statue or the valve in question spatially coincides with the material from which it is constructed (lump of clay, piece of metal machinery), it is not identical with the latter on the Kant-Sellars construal of identity since it does not have the same modal properties with it: the statue has the modal property that it would be destroyed if we reshaped it into a sphere, while the lump of clay does not have this property since it could well survive the radical reshaping of the statue. Likewise, the piece of metal machinery would survive if it were removed from the environment in which it plays the functional role of a valve, since it could very well play a different functional role in a different environment, whereas the valve obviously would not survive a change of circumstance that made it the case that it no longer functions as a valve. This means that the statue and the lump of clay, or likewise, the valve and the piece of metal machinery, are not identical. Of course, it does not follow from this that these (different) items are not intimately related. In fact, they do stand in an intimate relation, which may be called "material constitution." But a relation of material constitution is not an *identity* relation, and this is the radical consequence of the Kant-Sellars thesis if it is applied to identity claims.

Many philosophically important relations are analogous to the statue-clay relation, including the relation between psychological states, which are normatively identified and individuated, and the non-normative material mechanisms and processes in which those psychological states are realized. More importantly, the relation between the objects referred to in the manifest image and those referred to in the scientific image, being as it is analogous to the statue-clay relation, cannot be that of identity (though it can be that of material constitution). This is because the descriptive concepts of the manifest image come with criteria of identity and individuation that essentially involve counterfactual conditionals formulated in other descriptive terms that *also* belong to the manifest image. As Brandom puts this point:

> the criteria of identity and individuation for such descriptive sortals articulating denizens of the manifest image as 'credenza', 'violin', 'yawl', 'rocker panel', 'shrub', 'mortgage lien', 'stock market crash', 'ciborium', 'frock', 'crepe', 'tragedy' [cannot be specified] without appealing to their criteria of application, which are in turn couched in a plethora of further manifest-image descriptive vocabulary drawn from the relevant domains: interior decorating, musical, nautical, automotive, legal, and so on.
>
> (Brandom 2015, 77–78)

Subjunctive conditionals specified in manifest-image kind terms will not match those specified in their alleged successor scientific-image terms, since manifest-image kinds are identified and individuated functionally by their relations to things of other such (manifest-image) functional kinds, in complex systems articulated by social norms. And this means that, by Sellars' own lights (i.e., by the Kant-Sellars thesis about identity) the relation of manifest-image kinds to their "successor" kinds, specified in the vocabulary of an eventual natural science, may be that of material constitution, but it will certainly not be that of *identity*.[3] There are no true identity claims relating the descriptive terms (and corresponding objects) of the manifest image to successor descriptive terms drawn from the vocabulary of the scientific image.

4. The Relation between the Objects of the Manifest Image and the Objects of the Scientific Image Is Not One of Identity

I shall now argue that Brandom's Kant-Sellars thesis about modality, as applied to identity claims, does not undermine the Sellarsian *scientia mensura* principle, because the latter does not construe the relation between the objects of the manifest image and the objects of the scientific image as one of identity. The descriptive resources of the scientific image are superior to those of the manifest image in matters of ontology because they provide better *explanations* of the world and of our place in it than do those of the scientific image. This is because, according to Sellars, empirical generalizations couched in manifest-image "lifeworld" kind terms are never fully lawful, but are in fact invariably unstable. And what looks to be mere random (unpredicted) variation among cases in the manifest-image empirical generalizations is revealed to be lawful when seen from the perspective of scientific-image "successor" kind terms (thereby leading to the formulation of new generalizations that would remain opaque to us if we insisted on remaining to the manifest-image level) (Sellars 1961). Thus, the relation between the descriptive resources of the two images is an explanatory relation. However, the explanatory relation in question is not one of identity. It is true that, according to Sellars, the (ideal) scientific image ultimately explains in resolutely *non-normative* terms the normatively-laden and functionally-individuated phenomena, whose conditions of identity and individuation are couched in manifest-image descriptive and explanatory vocabulary. But notice that this explanation proceeds by *eliminating* manifest-image normative descriptions and explanations from the inventory of the basic ontological categories (categorial concepts) with which the spatiotemporal world is described and explained in full generality and maximal specificity. And scientific-image predicates do this not by having an *identity* relation with manifest-image predicates but by *changing* their modal implications (i.e., by changing their very meaning), thereby reconceptualizing or,

even, *recategorizing* the latter (in their descriptive and explanatory dimension) (Sellars 1961, 1963b).

Brandom's Kant-Sellars argument could indeed have force against more traditional, deductivist views of the explanatory relation between the manifest and the scientific image, according to which explanation should be understood in terms of derivation or logical deduction (i.e., as involving relations of derivability or conceptual entailment between the explanans and the explanandum) (Hempel 1965). Notice that this presupposes that the domain of the explanans shares the *same* meaning (conceptual content) with that of the explanandum. However, Sellars firmly rejects this view of explanation. As was mentioned above, scientific-image phenomena explain manifest-image phenomena not by having an identity relation with them but by having *different* modal properties than the latter. That is to say, scientific image kind terms explain manifest image terms by changing their very meaning, thereby reconceptualising—or even recategorizing—the latter in their descriptive and explanatory dimension on the basis of analogically construed successor or "counterpart" concepts. In this way, phenomena described and explained in manifest-image terms are reconceived as being appearances of an underlying reality that explains *why* the former are appearances and also why they *seem* to—or what amounts to the same thing, why *for all practical purposes* they indeed do—have real worldly efficacy. In other words, reconceiving manifest-image phenomena as appearances of an underlying reality described in scientific-image terms explains why the former are, at best, approximations of the truth (in the dimension of describing and explaining the world) yet strictly speaking false.[4] The fact that manifest-image phenomena are explained in scientific-image terms implies that phenomena described in manifest-image terms do not really exist, in the sense that they can be considered as implicated in causally efficacious phenomena (i.e., as providing ideally adequate explanations) only if they are *recategorized* and ultimately understood in *different*, *scientific*-image terms. Manifest-image objects do not really exist, except as reconceived *scientific* objects—or better, scientifically described "absolute processes"—namely as something radically, indeed categorially, *different* from manifest-image objects.

Now, it could be objected here, on Brandom's behalf, that his Kant-Sellars thesis does not have force only against a deductivist construal of the explanatory relation between the manifest and the scientific image. This is because Brandom construes the identity relation between the manifest and the scientific image as belonging in the realm of reference and not in that of sense; hence, he could equally object to a view according to which scientific-image explanations do indeed change the meaning but retain the reference or denotation of manifest-image terms. Yet, notice that, according to Sellars, the change of meaning in scientific-image explanations of manifest-image phenomena works by *eliminating* manifest-image objects, which means that the successor scientific-image concepts cannot possibly

be understood as retaining "the same" reference as their manifest-image predecessors.

It is true that Sellars is not always as clear as he should be as regards this issue. For instance, Sellars (1967, 138) seems to construe the relation between predecessor manifest-image terms and successor scientific-image terms as referential, in the extensional sense of the term. However, I take it that Sellars is indeed in a position to argue that the categorial transition from the manifest to the scientific-image framework goes hand in hand with a change of *reference* and not only of sense. Recall, for example, that, for Sellars, although reference is of course to be distinguished from sense, it is not a word-world relation—that is, it does not assert relations between linguistic and extra-linguistic items[5]—but instead belongs firmly within the 'order of *signification*' (i.e., in the space of reasons) (Sellars 1967, 82–87). In other words, for Sellars, extensions are a variety of sense, indeed a limiting case of intensions. Concepts such as 'reference' or 'denotation' cannot be understood apart from intensions and have normative import[6] (Sellars 1967, 77, 86). This means that radical differences in categorial role between concepts across frameworks cannot leave their reference unaffected. Now, as was mentioned above, manifest-image objects do not really exist, except as reconceived scientifically described "absolute processes." Yet, the categorial structure of scientific-image absolute processes (i.e., the successor concept of "manifest-image physical object" in the ideal scientific image) is so different from that of manifest-image objects as to block the possibility of any reference continuity between the successor scientific-image concepts in question and their manifest-image predecessors. For example, in a world of absolute processes the primary "objects" are no longer substances (things) that endure through change, belong to kinds, and have conditional (e.g., causal/dispositional) properties as criteria for belonging to the latter (Sellars 1957, 263–64). Note also, that, according to Sellars, the concept of 'absolute process' does not stand in a part-whole relation to objects *at all*. That is to say, processes are not "wholes" of which objects are "parts." Instead, processes will turn out to be, not further objects themselves, but virtual classes of objects (Sellars 1967, 149–50).

5. Sellars, Eliminativism, and Revisionism

But was Sellars so radical as to believe the manifest-image objects and their properties cannot provide really adequate descriptions and explanations of the world (all things considered) and hence do not really exist as such? Indeed he was, but with a twist. Consider, for example, the following passages:

> *Speaking as a philosopher*, I am quite prepared to say that the common sense world of [perceptible] physical objects in Space and Time is unreal-that is, that there are no such things.
>
> (Sellars 1997, §41)

> On the view I propose, the assertion that the micro-entities of physical theory really exist goes hand in hand with the assertion that the macro-entities of the perceptible world [physical objects in space and time] do not really exist. . . . The perceptual world is phenomenal in something like the Kantian sense, the key difference being that the real or 'noumenal' world which supports the 'world of appearances' is not a metaphysical world of unknowable things but simply the world as construed by scientific theory.
>
> (Sellars 1963c, 96–97)

> According to the view I am proposing, correspondence rules [linking observational and theoretical predicates] would appear in the material mode as statements to the effect that the objects of the observational framework [of physical objects in space and time] *do not really exist—there are really no such things*. They envisage the *abandonment* of a sense and its denotation [of observational predicates].
>
> (Sellars 1961, 77)

> A consistent scientific realist must hold that the world of everyday experience is a phenomenal world in the Kantian sense, existing only in as the contents of actual and obtainable conceptual representings, the obtainability of which is explained . . . by scientific objects.
>
> (Sellars 1967, 173)

Yet, at the same time, Sellars is quick to point out that to say that perceptible physical objects in space and time do not really exist—that physical objects are not really colored, are not located in space, do not endure through time—is to make a point (from a *transcendental* point of view) *about* the manifest-image framework *as a whole*, not *in* it. It does not amount to a rejection of an empirical belief formulated within the framework. That is, it is not to say that the sentence "physical objects have colors" expresses an empirical proposition, which, though widely believed by common sense, has been shown by science to be false. As long as we are in the manifest-image framework, we evaluate statements about physical objects, their properties, and their relations to perceivers in terms of criteria provided by this very framework. The manifest-image framework is a tightly interconnected whole, and the attempt to replace it *piecemeal* by fragments of the scientific image is a cause of philosophical confusion.

According to Sellars, to say that the phenomena described and explained in manifest-image kind terms do not really exist is an expression of a wholesale (*not* piecemeal) rejection of the very framework itself, in its descriptive and explanatory dimension, in favor of another built around different, though not unrelated, categories—namely, the ideal scientific image. It is also to say that science is making available a more adequate framework of entities, principles, and laws (in the dimension of describing and explaining

the world), which, in principle at least, could serve all the perceptual-inferential-practical functions of the framework that we employ in everyday life (Sellars 1963c, 97; 1997, §42). Again, this is not to say that there is good reason for adopting this envisaged new scientific framework in *practice*. Indeed there are sound *methodological* reasons for *not* teaching ourselves to respond to perceptible situations in terms of constructs in the language of the scientific-image. For as Sellars puts it:

> While this could, in principle, be done, the scientific quest is not yet over, and even granting that the main outlines are blocked in, the framework of physical objects in space and time, shaped over millennia of social evolution, provides when accompanied by correct philosophical commentary, a firm base of operations with which to correlate the developing structure of scientific theory, refusing to embrace any stage without reserve as our very way of perceiving the world, not because it would not be a *better* way, but because the better is the enemy of the best.
>
> (Sellars 1963c, 97)

Moreover, importantly, we should be reminded of the fact that, for Sellars, *it is only in the dimension of describing and explaining* the world that we can say of manifest-image phenomena that they do not really exist. And this view, as such, is compatible with the thesis that these 'phenomena' *do* really exist, and irreducibly so, if they are considered as being *functionally* (as opposed to being 'materially' or 'qualitatively) individuated in terms of the *normative* (rather than explanatory) role they play in the context of a whole behavioral economy of human organisms living and interacting with other such organisms. In *this* sense, that is, considered as normatively individuated functional roles, manifest-image concepts and categories (e.g., the concept of a person who believes, intends, suffers, thinks, acts, desires, has certain capacities or abilities or related concepts thereof such as that of an 'artifact', 'social institution', etc.) remain absolutely intact even in the ideal scientific image of man-in-the-world. On the other hand, in the ideal scientific image, there would be a radical change in the *"material" (or "qualitative") content* of those, and indeed of all normatively laden, concepts[7] (see also O'Shea 2011, 330–37).

Indeed, the view developed in section 4 can be further refined by highlighting the fact that while Sellars can be considered as an eliminativist about manifest-image kinds from a *transcendental* point of view, which attempts to achieve a view *across* different conceptual frameworks considered as *wholes*, he can equally be understood as putting forward an "identity-through-difference" *revisionary* view about manifest-image kind terms. This latter view can be justified if manifest-image terms are understood in their *historical-developmental* dimension, namely as *historically evolving conceptual roles* that express the gradual "contamination" of the manifest-image framework with scientific-image kind terms. Note that this

is a *piecemeal revisionary* process leading gradually from the descriptive and explanatory categories of the former to those of the latter.

Moreover, interestingly, from this point of view (i.e., that of our *developing/evolving epistemic practices*), we can make sense of *both* of the following, seemingly opposed claims regarding the conditions of identity and individuation of manifest-image objects: 1) Our conception of the *explanatory* adequacy of manifest-image predicates may well undergo radical *change*, in which case the conditions of identity and individuation of manifest-image objects will indeed be radically modified. 2) There is *another* dimension of the use/function of manifest-image terms, namely their "abstract" *normative* functional *role*, in which the conditions of identity and individuation of manifest-image objects—this time understood in a suitably *'schematic'* sense[8]—do *not* undergo radical change throughout radical changes in the explanatory ("contentual") function of manifest-image predicates. Another way to make these points would be to say that predecessor manifest-image concepts are *generically similar* while at the same time *specifically different* from the successor scientific-image concepts (see also Sellars 1973).

Now consider again the basic point made above. It was suggested that an "identity-through-difference" revisionary view about manifest-image kind terms, which emphasizes their historical-developmental dimension, enables us to hold both of the following views simultaneously: 1) There is a suitably schematic and "ideal" sense (related to normative function) in which the conditions of identity and individuation of manifest-image objects do not undergo radical change. 2) There is another "contentual" sense (related to explanatory function), in which the conditions of identity and individuation in question—and hence, the manifest-image predicates and objects—undergo radical change. Now, building on the above points, it might also be proposed that, for Sellars, those two points are "dialectically" related in the sense that there is a rationale (i.e., accommodation of explanatory anomalies) that justifies this radical change as *necessary* for the very *preservation* of the normative (regulative) ideal[9] of the very descriptive and explanatory practices in which *manifest*-image kind terms themselves are embedded (see also Garfield 2012, 106–09). And to switch again to the transcendental point of view, we might even say that, for Sellars, somewhat paradoxically, it is precisely the preservation of the normative (regulative) ideal built in the descriptive and explanatory dimension of the use of *manifest*-image kind terms that necessitates their eventual *abandonment* in favor of (ideal-"Peirceian") *scientific*-image kind terms.

6. Concluding Remarks

In sum, it seems that Sellars is in a position to accommodate Brandom's Kant-Sellars thesis about modality and identity without abandoning his *scientia mensura* principle. Brandom's Kant-Sellars thesis is indeed in tension with two ways of understanding of this principle: 1) the view that the

explanatory relation between scientific-image and manifest-image objects should be conceived as a relation of conceptual entailment, and 2) the view according to which scientific-image explanations change the meaning, but retain the reference or 'denotation' of manifest-image terms. Yet, Sellars firmly criticizes and explicitly rejects (1), and he would equally reject (2) on the grounds that since the change of meaning in scientific-image explanations of manifest-image phenomena works by eliminating manifest-image objects, the successor scientific-image concepts, themselves construed in non-objectual terms (absolute processes), cannot possibly be understood as retaining "the same" reference as their manifest-image 'predecessors'.

Moreover, although there is a sense in which manifest-image kind terms are abandoned for explanatory reasons in favor scientific-image kind terms, it does not follow from this that the normative aspect of the manifest image is eliminated. This is because, according to Sellars, 1) the correct *philosophical* analysis of normative discourse in the manifest image shows that its function is not ultimately *explanatory* (though in its *actual* 'everyday' use within the 'lifeworld' normative discourse *does* have an explanatory dimension), and 2) in this *non*-explanatory sense (i.e., as something that specifies certain functions or roles that various empirical [linguistic or non-linguistic] items play, which serve as standards for assessing and prescribing their actual 'behavior'), the normativity of the manifest-image framework proves to be strictly *irreducible* to non-normative scientific-image terms. Hence, Sellars cannot be accused of eliminating the normativity of the manifest image *as such* even if there is indeed an (explanatory) sense in which normative phenomena "really exist" only if they are reconceived in non-normative terms.

Notes

1 By 'labelling' Sellars means discriminating, in the sense of responding differentially.

2 As Brandom puts this point: "Part of taking an inference to be materially good is having a view about which possible additional collateral premises or auxiliary hypotheses would, and which would not, infirm it. Chestnut trees produce chestnuts—unless they are immature of blighted. Dry, well made matches strike—unless there is no oxygen" (Brandom 2015, 141–42). Importantly, this means that material inference is in general non-monotonic. That is, the inference from p to q may be materially good, even though the inference from $p\&r$ to q is not (Brandom 2015, 163–65).

3 Indeed, as Brandom points out in this connection, a particularly interesting consequence of the above considerations is that "*nothing* is identical to the mereological sum of things of other kinds (e.g. fundamental particles). This is . . . because mereological sums are indifferent to the spatial rearrangement of their parts-and that is not true of things of *any* of the [manifest-image] kinds. Nor are they identical to specifiable spatiotemporal constellations of their parts or particles. For not even a wildly disjunctive specification in the language of physics will underwrite the right subjunctive conditionals to agree in criteria of application and identity with those determined by the manifest-image kinds. For even if one could say, holding a great deal constant, what arrangements of particles *would* count as a stock-market crash or a mortgage lien, *what* one would have

to hold constant to do so would itself have to be specified at least in part using other manifest-image sortals and descriptive predicates" (Brandom 2015, 78).

4 Recall that approximate truth is a species of falsehood.

5 It might be objected here that even if reference is not a word-world relation it does not follow from this that reference is not a relation between linguistic and extra-linguistic items, but only that reference is a matter of *many* word-world relations (see e.g., Brandom 1994, 314). However, whatever the merits of this position, it is not Sellars' own view. According to Sellars, "semantical statements of the Tarski-Carnap variety [including 'x' stands for y, 'x' denotes y, 'x' is true of y] do not assert relations between linguistic and non-linguistic items, though, in the case of expressions which stand for senses which are intensions, it will also be true (and necessarily so) that these expressions are involved in semantical uniformities (actual or potential) with the appropriate extra-linguistic items" (Sellars 1967, 82).

6 For example, the concept of a word's *denoting Chicago* needs to be understood in terms of its impact on, say, a sentence's *being true iff Chicago is large*, which in turn needs to be understood as a requirement on the applicability of an *intentional* notion, namely the notion of a sentence's *expressing the thought that Chicago is large* (see also Shapiro 2013, 309).

7 Note also that in the manifest image, as it stands, the normatively individuated functional role of manifest-image kind terms is not strictly distinguished from its contentual/material aspect, and this entanglement results in what Sellars calls the "descriptive ontology of everyday life" (Sellars 1997, §41). Now, if we apply this methodological distinction to our ordinary "lifeworld" concepts (e.g., that of a 'table'), we can say that it is composed of two, inextricably entangled, elements: the table qua ready-to-hand object, essentially involving its character as an instrument we use to satisfy our purposes, and the table qua proper and common sensible properties (i.e., considered as an object of sensory and affective experience). And the purpose of breaking our ordinary 'lifeworld' concepts in those two parts is precisely in order to highlight the fact that this latter component can be radically transformed by successor scientific-image concepts, without necessarily eliminating the former.

8 For example, it might be thought that 'water' is picked out as "whatever stuff causes certain effects". In this case it would be a schematic (manifest-image) sortal which would really be a placeholder for less schematic (scientific-image) sortals. And the latter, less schematic scientific-image sortals, in turn, would be the expression of the normative ideals built into our manifest-image concept of 'water'—i.e. of what ideally *ought* to be the case for something to count as 'water'. Note that the sortals that identify and individuate intentional states and episodes are similarly schematic.

9 This regulative ideal can be understood as the demand for a systematic, unified, and stratified body of empirical knowledge capable of yielding arbitrarily general pictures of the world and our place in it at an arbitrary level of specificity in detail.

References and Further Reading

Brandom, Robert. 1994. *Making It Explicit*. Cambridge, MA: Harvard University Press.

Brandom, Robert. 2015. *From Empiricism to Expressivism: Brandom Reads Sellars*. Cambridge, MA: Harvard University Press.

deVries, Willem. 2005. *Wilfrid Sellars*. Chesham: Acumen Publishing.

deVries, Willem. 2012. "Ontology and the Completeness of Sellars' Two Images." *Humana.Mente* 21: 1–18.

Garfield, Jay. 2012. "Sellarsian Synopsis: Integrating the Images." *Humana.Mente* 21: 103–121.

Hempel, Carl. 1965. *Aspects of Scientific Explanation*. New York: Free Press.

O'Shea, James. 2011. "How to be a Kantian and a Naturalist about Human Knowledge: Sellars' Middle Way." *Journal of Philosophical Research* 36: 327–359.

Sellars, Wilfrid. 1957. "Counterfactuals, Disposition and the Causal Modalities." In *Minnesota Studies in the Philosophy of Science*, vol. II, edited by Herbert Feigl, Michael Scriven and Grover Maxwell, 225–308. Minneapolis, MN: University of Minnesota Press.

Sellars, Wilfrid. 1961. "The Language of Theories" In *Current Issues in the Philosophy of Science*, edited by Herbert Feigl and Grover Maxwell, 57–77. New York: Holt, Rinehart & Winston.

Sellars, Wilfrid. 1963a. "Philosophy and the Scientific Image of Man." In *Science, Perception and Reality*, edited by Wilfrid Sellars, 1–40. London: Routledge & Kegan Paul.

Sellars, Wilfrid. 1963b. "Theoretical Explanation." In *Philosophy of Science: The Delaware Seminar*, Vol. II, edited by Bernard Baumrin, 61–78. New York: John Wiley.

Sellars, Wilfrid. 1963c. "Phenomenalism." In *Science, Perception and Reality*, edited by Wilfrid Sellars, 60–105. London: Routledge & Kegan Paul.

Sellars, Wilfrid. 1967. *Science and Metaphysics: Variations on Kantian Themes*. Atascadero, CA: Ridgeview Publishing Company.

Sellars, Wilfrid. 1973. "Conceptual Change." In *Conceptual Change*, edited by Patrick Maynard and Glenn Pearce, 77–93. Dordrecht: D. Reidel Publishing.

Sellars, Wilfrid. 1997. *Empiricism and the Philosophy of Mind*. Cambridge, MA: Harvard University Press.

Shapiro, Lionel. 2013. "Intentional Relations and the Sideways-on View: On McDowell's Critique of Sellars". *European Journal of Philosophy* 21(2): 300–319.

12 Categories and Noumena

Two Kantian Axes of Sellars' Thought

Robert B. Brandom

PART I: ON THE WAY TO A PRAGMATIST THEORY OF THE CATEGORIES

1. Introduction

Several decades ago, Richard Rorty suggested that philosophical admirers of Wilfrid Sellars could be divided into two schools, defined by which of two famous passages from his masterwork *Empiricism and the Philosophy of Mind* are taken to express his most important insight:

> In the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not.
>
> (§41)

or

> [In] characterizing an episode or a state as that of *knowing,* we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says.
>
> (§36)[1]

The first passage, often called the "*scientia mensura,*" expresses a kind of scientific naturalism. Its opening qualification is important: there are other discursive and cognitive activities besides describing and explaining. The second passage says that characterizing something as a knowing is one of them. And indeed, Sellars means that in characterizing something even as a believing or a believable, as conceptually contentful at all, one is doing something other than describing it. One is placing the item in a normative space articulated by relations of what is a reason for what. Meaning, for him, is a normative phenomenon that does not fall within the descriptive realm over which natural science is authoritative.

Rorty called those impressed by the scientific naturalism epitomized in the *scientia mensura* "right-wing Sellarsians" and those impressed by the normative nonnaturalism about semantics expressed in the other passage "left-wing Sellarsians." Acknowledging the antecedents of this usage, he used to express the hope that right-wing and left-wing Sellarsians would be able to discuss their disagreements more amicably and irenically than did the right-wing and left-wing Hegelians, who, as he put it, "eventually sorted out their differences at a six-month-long seminar called 'the Battle of Stalingrad.'" According to this botanization, I am, like my teacher Rorty and my colleague John McDowell, a left-wing Sellarsian, by contrast to such eminent and admirable right-wing Sellarsians as Ruth Millikan, Jay Rosenberg, and Paul Churchland.

While I think Rorty's way of dividing things up is helpful, I want here to explore a different perspective on some of the same issues. I, too, will focus on two big ideas that orient Sellars' thought. I also want to say that one of them is a good idea, and the other one on the whole is a bad idea—a structure that is in common between those who would self-identify as either right- or left-wing Sellarsians. And the one I want to reject is near and dear to the heart of the right-wing. But I want, first, to situate the ideas I'll consider in the context of Sellars' neo-Kantianism: they are his ways of working out central ideas of Kant's. Specifically, they are what Sellars makes of two fundamental ideas that are at the center of Kant's transcendental idealism: the metaconcept of *categories,* or *pure concepts of the understanding,* and the distinction between *phenomena* and *noumena*. The latter is a version of the distinction between appearance and reality, not in a light epistemological sense, but in the ontologically weighty sense that is given voice by the *scientia mensura*. I cannot say that these fall under the headings, respectively, of What Is Living and What Is Dead in Sellars' thought, since the sort of scientific naturalism he uses to interpret Kant's phenomena/noumena distinction is undoubtedly very widespread and influential in contemporary Anglophone philosophy. My aim here is threefold: to explain what I take it Sellars makes of these Kantian ideas, why I think the first line of thought is more promising than the second, and the way forward from each that seems to me most worth developing.

When asked what he hoped the effect of his work might be, Sellars said he would be happy if it helped usher analytic philosophy from its Humean into its Kantian phase. (*A propos* of this remark, Rorty also said, not without justice, that in these terms my own work could be seen as an effort to help clear the way from analytic philosophy's incipient Kantian phase to an eventual Hegelian one.)[2] Sellars tells us that his reading of Kant lies at the center of his work. He used that theme to structure his John Locke lectures, to the point of devoting the first lecture to presenting a version of the Transcendental Aesthetic with which Kant opens the *Critique of Pure Reason*. Those lectures, published as *Science and Metaphysics: Variations on Kantian Themes,* are Sellars' only book-length, systematic exposition

of his views during his crucial middle period. The development of Kantian themes is not only self-consciously used to give that book its distinctive shape but also implicitly determines the contours of Sellars' work as a whole. I think the best way to think about Sellars' work is as a continuation of the neo-Kantian tradition. In particular, I think he is the figure we should look to today in seeking an appropriation of Kant's theoretical philosophy that might be as fruitful as the appropriation of Kant's practical philosophy that Rawls initiated. On the theoretical side, Sellars was the greatest neo-Kantian philosopher of his generation.[3]

In fact, the most prominent neo-Kantians of the previous generation, C. I. Lewis and Rudolf Carnap, were among the most immediate influences on Sellars' thought. Kant was the door through which Lewis found philosophy and later, the common root to which he reverted in his attempt to reconcile what seemed right to him about the apparently antithetical views of his teachers, William James and Josiah Royce. (Had he instead been trying to synthesize Royce with Dewey, instead of James, he would have fetched up at Hegel.) In his 1929 *Mind and the World Order,* Lewis introduced as a central technical conception the notion of the sensory "Given," which Sellars would famously use (characteristically, without mentioning Lewis by name) as the paradigm of what he in *EPM* called the "Myth of the Given." (Indeed, shortly after his 1946 *An Analysis of Knowledge and Valuation,* which Sellars also clearly has in mind in *EPM,* Lewis wrote a piece addressing the question "Is the Givenness of the Given Given?" His answer was No: It is a necessary postulate of high philosophical theory, which dictates that without a sensory Given, empirical knowledge would be impossible.)

Sellars modeled his own Kantian "metalinguistic" treatments of modality and the ontological status of universals explicitly on ideas of Carnap. Although, like Lewis, Carnap is not explicitly mentioned in *EPM,* his presence is registered for the philosophical cognoscenti Sellars took himself to be addressing there by the use of the Carnapian term "protocol sentence" (as well as Schlick's "Konstatierung") for noninferential observations. Unlike Lewis, Carnap actually stood in the line of inheritance of classical 19th-century German neo-Kantianism. His teacher, Bruno Bauch, was (like Heidegger) a student of Heinrich Rickert in Freiburg—who, with the older Wilhelm Windelband, led the Southwest or Baden neo-Kantian school. In spite of these antecedents, Bauch was in many ways closer to the Marburg neo-Kantians, Hermann Cohen and Paul Natorp, in reading Kant as first and foremost a philosopher of the natural sciences, mathematics, and logic. I suppose that if one had asked Carnap in what way his own work could be seen as a continuation of the neo-Kantian tradition of his teacher, he would first have identified with this Marburg neo-Kantian understanding of Kant, and then pointed to the *logical* element of his logical empiricism—itself a development of the path-breaking work of Frege, Bauch's friend and colleague at Jena when Carnap studied with both there—as giving a precise and modern form to the conceptual element in empirical knowledge, which

deserved to be seen as a worthy successor to Kant's own version of the conceptual.

If Lewis and Carnap do not immediately spring to mind as neo-Kantians, that is because each of them gave Kant an empiricist twist, which Sellars was concerned to undo. If you thought that Kant thought the classical empiricists' Cartesian understanding of the sensory contribution to knowledge was pretty much all right, and just needed to be supplemented by an account of the independent contribution made by a conceptual element, you might well respond to the development of the new 20th-century logic with a version of Kant that looks like Lewis's *Mind and the World Order* and *An Analysis of Knowledge and Valuation,* and Carnap's *Aufbau* (and for that matter, Nelson Goodman's *Structure of Appearance*). That assumption about Kant's understanding of the role played by sense experience in empirical knowledge is exactly what Sellars challenges in *EPM*.

One of the consequences of his doing that is to make visible the neo-Kantian strand in analytic philosophy that Lewis and Carnap each, in his own way, represented—and which Sellars and in our own time, John McDowell, further developed. Quine was a student of both Lewis and Carnap, and the Kantian element of the common empiricism he found congenial in their thought for him drops out entirely—even though the logic remains. His Lewis and his Carnap are much more congenial to a narrative of the history of analytic philosophy initiated by Bertrand Russell and G. E. Moore, according to which the movement is given its characteristic defining shape as a recoil from Hegel (seen through the lenses of the British Idealism of the waning years of the 19th century). They understood enough about the Kantian basis of Hegel's thought to know that a *holus bolus* rejection of Hegel required a diagnosis of the idealist rot as having set in already with Kant. This narrative does pick out one current in the analytic river—indeed, the one that makes necessary the reappropriation of the metaconceptual resources of Kant's theoretical philosophy in the late 20th and early 21st centuries. But it was never the whole story.[4] The neo-Kantian tradition comprising Lewis, Carnap, and Sellars can be thought of as an undercurrent, somewhat occluded from view by the empiricist surface.

2. Categories in Kant

Many Kantian themes run through Sellars' philosophy. I am going to focus on two master-ideas, each of which orients and ties together a number of otherwise apparently disparate aspects of his work. The first is the idea that besides concepts whose characteristic expressive job it is to describe and explain empirical goings-on, there are concepts whose characteristic expressive job it is to make explicit necessary structural features of the discursive framework within which alone description and explanation are possible. Failing to acknowledge and appreciate this crucial difference between the expressive roles different bits of vocabulary play is a perennial source of

distinctively philosophical misunderstanding. In particular, Sellars thinks, attempting to understand concepts doing the second, framework-explicating sort of work on the model of those whose proper use is in empirical description and explanation is a fount of metaphysical and semantic confusion.[5] Among the vocabularies that play the second sort of role, Sellars includes *modal* vocabulary (not only the alethic, but also the deontic species), *semantic* vocabulary, *intentional* vocabulary, and *ontological-categorial* vocabulary (such as 'proposition', 'property' or 'universal', and 'object' or 'particular'). It is a mistake, he thinks, to understand the use of any of these sorts of vocabulary as fact-stating in the narrow sense that assimilates it to *describing* how the world is. It is a corresponding mistake to recoil from the metaphysical peculiarity and extravagance of the kinds of facts one must postulate in order to understand statements couched in these vocabularies as fact-stating in the narrow sense (e.g., normative facts, semantic facts, conditional facts, facts about abstract universals) by denying that such statements are legitimate, or even that they can be true. (Though to say that they are true is not, for Sellars, to describe them.) Both mistakes (the dogmatic metaphysical and the skeptical), though opposed to one another, stem from the common root of the *descriptivist fallacy.* That is the failure to see that some perfectly legitimate concepts do not play a narrowly descriptive role, but rather a different, explicative one with respect to the practices of description and explanation. Following Carnap, Sellars instead analyzes the use of all these kinds of vocabulary as, each in its own distinctive way, "covertly metalinguistic."

In opposing a Procrustean descriptivism about the expressive roles locutions can play, Sellars makes common cause with the later Wittgenstein. For Wittgenstein, too, devotes a good deal of effort and attention to warning us of the dangers of being in thrall to ("bewitched by") a descriptivist picture. We must not simply assume that the job of all declarative sentences is to state facts ("I am in pain," "It is a fact that . . ."), that the job of all singular terms is to pick out objects ("*I* think . . ."), and so on. In addition to tools for attaching, detaching, and in general reshaping material objects (hammer and nails, saws, draw-knives, . . .), the carpenter's tools also include plans, a foot-rule, level, pencil, and tool belt. So, too, with discursive expressive ˢtoolsˢ. Wittgenstein's expressive pluralism (language as a motley) certainly involves endorsement of the anti-descriptivism Sellars epitomizes by saying

> [O]nce the tautology 'The world is described by descriptive concepts' is freed from the idea that the business of all non-logical concepts is to describe, the way is clear to an *ungrudging* recognition that many expressions which empiricists have relegated to second-class citizenship in discourse are not *inferior,* just *different*.[6]

But Sellars differs from Wittgenstein in characterizing at least a broad class of nondescriptive vocabularies as playing generically the *same* expressive

role. They are broadly metalinguistic locutions expressing necessary features of the framework of discursive practices that make description (and—so—explanation) possible. Of this broad binary distinction of expressive roles, with ordinary empirical descriptive vocabulary on one side and a whole range of apparently disparate vocabularies going into another class as "metalinguistic," there is, I think, no trace in Wittgenstein.[7]

The division of expressive roles I am claiming for Sellars binds together modal, semantic, intentional, and ontological-categorial vocabulary in opposition to empirical descriptive vocabularies and traces back to Kant's idea of "pure concepts of the understanding," or categories, which play quite a different expressive role from that of ordinary empirical descriptive concepts. The expressive role of pure concepts is, roughly, to make explicit what is implicit in the use of ground-level concepts: the conditions under which alone it is possible to apply them, which is to say, use them to make judgments. Though very differently conceived, Kant's distinction is in turn rooted in the epistemological difference Hume notices and elaborates between ordinary empirical descriptive concepts and concepts expressing lawful causal-explanatory connections between them. Hume, of course, drew skeptical conclusions from the observation that claims formulated in terms of the latter sort of concept could not be justified by the same sort of means used to justify claims formulated in terms of empirical descriptive concepts.

Kant, however, looks at Newton's formulation of the best empirical knowledge of his day and sees that the newly introduced concepts of *force* and *mass* are not intelligible apart from the laws that relate them. If we give up the claim that F equals m*a then we do not mean *force* and *mass,* but are using some at least slightly different concepts. (Galileo's geometrical version of the—late medieval—observable concept of *acceleration is* antecedently intelligible.) This leads Kant to two of his deepest and most characteristic metaconceptual innovations: thinking of statements of laws formulated using alethic modal concepts as making explicit rules for reasoning with ordinary empirical descriptive concepts, and understanding the contents of such concepts as articulated by those rules of reasoning with them.

This line of thought starts by revealing the semantic presuppositions of Hume's epistemological arguments. For Hume assumes that the contents of ordinary empirical descriptive concepts are intelligible antecedently to and independently of taking them to stand to one another in rule-governed inferential relations of the sort made explicit by modal concepts. Rejecting that semantic atomism then emerges as a way of denying the intelligibility of the predicament Hume professes to find himself in: understanding ordinary empirical descriptive concepts perfectly well, but getting no grip thereby on the laws expressed by subjunctively robust rules relating them. Even though Kant took it that Hume's skeptical epistemological argument rested on a semantic mistake, from his point of view Hume's investigation had uncovered a crucial *semantic* difference between the expressive roles of

different kinds of concepts. Once his attention had been directed to them, he set himself the task of explaining what was special about these *non*descriptive concepts.

Two features of Kant's account of the expressive role distinctive of the special class of concepts to which Hume had directed his attention are of particular importance for the story I am telling here. They are *categorial* concepts, and they are *pure* concepts. To say that they are 'categorial' in this context means that they make explicit aspects of the *form* of the conceptual as such. For Kant concepts are functions of judgment—that is, they are to be understood in terms of their role in judging. Categorial concepts express structural features of empirical descriptive judgments. What they make explicit is implicit in the capacity to make any judgments at all. This is what I meant when I said above that rather than describing how the world is, the expressive job of these concepts is to make explicit necessary features of the framework of discursive practices within which it is possible to describe how the world is. The paradigm here is the alethic modal concepts that articulate the subjunctively robust consequential relations among descriptive concepts.[8] It is those relations that make possible *explanations* of why one description applies because another does. That force *necessarily* equals the product of mass and acceleration means that one can explain the specific acceleration of a given mass by describing the force that was applied to it. (Of course, Kant also thinks that in articulating the structure of the judgeable as such, these concepts *thereby* articulate the structure of what is empirically *real:* the structure of *nature,* of the *objective world.* But this core thesis of his understanding of empirical realism within transcendental idealism is an optional additional claim, not entailed by the identification of a distinctive class of concepts as categories of the understanding.)

To say that these concepts are 'pure' is to say that they are available to concept-users (judgers = those who can understand, since for Kant the understanding is the faculty of judgment) *a priori.*[9] Since what they express is implicit in any and every use of concepts to make empirical judgments, there is no *particular* such concept one must have or judgment one must make in order to be able to deploy the pure concepts of the understanding. To say that judgers can grasp these pure concepts *a priori* is *not* to say that they are *immediate* in the Cartesian sense of nonrepresentational. Precisely not. The sort of self-consciousness (awareness of structural features of the discursive as such) they make possible is mediated by those pure concepts. What was right about the Cartesian idea of the immediacy of self-consciousness is rather that these mediating concepts are available to every thinker *a priori.* Their grasp does not require grasp or deployment of any *particular* ground-level empirical concepts but is *implicit* in the grasp or deployment of *any* such concepts. The way I will eventually recommend we think about this distinctive *a prioricity* is that in being able to deploy ordinary empirical descriptive concepts one already knows how to do everything one needs to know how to do in order to be able to deploy the concepts that play the

expressive role characteristic of concepts Kant picks out as "categorial" (as well as some that he does not).

3. Categories in Sellars

Sellars' development of Kant's idea of pure concepts of the understanding is articulated by two master-ideas. First, his successor metaconception comprises concepts that are in some broad sense *metalinguistic.*[10] In pursuing this line he follows Carnap, who, besides ground-level empirical descriptive vocabulary, allowed metalinguistic vocabulary as also legitimate in formal languages regimented to be perspicuous. Such metalinguistic vocabulary allows the formulation of explicit rules governing the use of descriptive locutions. Ontologically classifying terms such as 'object', 'property', and 'proposition' are "quasi-syntactical" metavocabulary corresponding to overtly syntactical expressions in a proper metalanguage such as 'singular term', 'predicate', and 'declarative sentence'. They are used to formulate "L-rules," which specify the structure of the language in which empirical descriptions are to be expressed.[11] Alethic modal vocabulary is used to formulate "P-rules," which specify rules for reasoning with particular empirically contentful descriptive vocabulary. Carnap's neo-Kantianism does not extend to embracing the metaconcept of *categories,* which he identifies with the excesses of transcendental idealism. But in the expressions Carnap classifies as overtly or covertly metalinguistic, Sellars sees the raw materials for a more thoroughly Kantian successor conception to the idea of pure categories of the understanding.

The second strand guiding Sellars' reconceptualization of Kantian categories is his *semantic inferentialist* approach to understanding the contents of descriptive concepts. Sellars picks up on Kant's rejection of the semantic atomism characteristic of both the British empiricism of Locke and Hume that Kant was reacting to and of the logical empiricism of Carnap that Sellars was reacting to.[12] The way he works out the anti-atomist lesson he learns from Kant is in terms of the essential contribution made to the contents of ordinary empirical descriptive concepts by the inferential connections among them appealed to in *explanations* of why some descriptions apply to something in terms of other descriptions that apply to it.

> Although describing and explaining (predicting, retrodicting, understanding) are *distinguishable,* they are also, in an important sense, *inseparable.* It is only because the expressions in terms of which we describe objects, even such basic expressions as words for perceptible characteristics of molar objects, locate these objects in a space of implications, that they describe at all, rather than merely label. The descriptive and explanatory resources of language advance hand in hand.[13]

This is a rich and suggestive passage. It is worth unpacking the claims it contains. It is framed by a distinction between a weaker notion, *labeling,* and a stronger one, *describing*. By 'labeling' Sellars means discriminating, in the sense of responding differentially. A linguistic expression is used as a label if its *whole* use is specified by the *circumstances* under which it is applied—the *antecedents* of its application. We might distinguish between three kinds of labels, depending on how we think of these circumstances or antecedents. First, one could look at what stimuli as a matter of fact elicit or in fact have elicited the response that is being understood as the application of a label. Second, one could look *dispositionally* at what stimuli *would* elicit the application of the label. Third, one could look at the circumstances in which the label is *appropriately* applied. What the three senses have in common is that they look only *up*stream, to the situations that have, would, or should prompt the use of the label. The first provides no constraint on future applications of the label—*que sera sera*—as familiar gerrymandering arguments about "going on in the same way" remind us. The second doesn't fund a notion of mistaken application. However one is disposed to apply the label is proper, as arguments summarized under the heading of "disjunctivitis" make clear. Only the third, normatively richer sense in which the semantics of a label consists in its circumstances of *appropriate* application (however the proprieties involved are understood) makes intelligible a notion of *mislabeling*.

Sellars wants to distinguish labeling in *all* of these senses from *describing*. The idea is that since labeling of any of these sorts looks only to the *circumstances* in which the label is, would be, or should be applied, expressions used with the semantics characteristic of labels address at most one of the two fundamental aspects of the use characteristic of descriptions. The rules for the use of labels tell us something about what is (or would be or should be) in effect so described, but say nothing at all about what it is described *as*. That, Sellars thinks, depends on the *consequences* of applying one description rather than another. The semantics of genuine descriptions must look downstream, as well as upstream. It is this additional feature of their use that distinguishes descriptions from labels. Here one might quibble verbally with Sellars' using 'label' and 'description' to describe expressions whose semantics depend on only one or on both of these dimensions of use. But it seems clear that a real semantic distinction is being marked.

Making a further move, Sellars understands those consequences of application of descriptions as essentially involving *inferential* connections to other descriptive concepts. This is what he means by saying that what distinguishes descriptions from labels is their situation in a "space of implications." We can think of these implications as specifying what other descriptions do, would, or should *follow from* the application of the initial, perhaps responsively elicited, description. As he is thinking of things, a description (correctly) applies to a range of things (for descriptive concepts used observationally, including those that are appropriately

noninferentially differentially responded to by applying the concept), which are described *by* it. And it describes them *as* something from which a further set of descriptions (correctly) follows. Crucially, these further descriptions can themselves involve applications of descriptive concepts that also have *non*inferential (observational) circumstances of application. Descriptive concepts that have *only* inferential circumstances of application he calls 'theoretical' concepts.

In the opening sentence of the passage, Sellars includes *understanding* as one of the phenomena he takes to be intricated with description in the way explaining is. Understanding a descriptive concept requires being able to place it in the "space of implications," partly in virtue of which it has the content that it does. This is in general a kind of knowing *how* rather than a kind of knowing *that:* being able to distinguish in practice the circumstances and consequences of application of the concept, when it is appropriately applied and what follows from so applying it. Grasping a concept in this sense is not an all-or-none thing. The ornithologist knows her way around inferentially in the vicinity of terms such as 'icterid' and 'passerine' much better than I do. A consequence of this way of understanding understanding is that one cannot grasp one concept without grasping many. This is Sellars' way of developing Kant's anti-atomist semantic insight.

Taking a further step (undertaking a commitment not yet obviously entailed by the ones attributed so far), Sellars also thinks that the inferences articulating the consequences of concepts used descriptively must always include *subjunctively robust* inferences. That is, the inferences making up the "space of implications" in virtue of which descriptive concepts have not only potentially atomistic circumstances of application but also nonatomistic relational consequences of application must extend to what other descriptions *would be* applicable if a given set of descriptions *were* applicable. For what Sellars means by 'explanation' is understanding the applicability of some descriptions as *explained by* the applicability of others according to just this kind of inference. This is, of course, just the sort of inferential connection that Hume's empiricist atomistic semantics for descriptive concepts, construing them as labels, could not underwrite. Sellars' conception of *descriptions,* as distinguished from labels, is his way of following out what he sees as Kant's anti-atomist semantic insight. *Modal* concepts make explicit these *necessary* inferential-consequential connections between descriptive concepts. They thereby perform the expressive role characteristic of Kantian categories: expressing essential features of the framework within which alone genuine description is possible.

All of this is meant to explicate what Sellars means by saying that "the descriptive and explanatory resources of language advance hand in hand." In addition to Kant's idea, Sellars here takes over Carnap's idea of understanding concepts whose paradigm is modal concepts as (in some sense) *metalinguistic.* The principal class of genuinely intelligible, nondefective, nondescriptive vocabulary Carnap allows in *The Logical Syntax of*

Language is syntactic metavocabulary and what he there calls "quasi-syntactical" vocabulary, which is covertly metalinguistic.[14] For Sellars, the *rules* modal vocabulary expresses are rules for deploying linguistic locutions. Their "rulishness" is their subjunctive robustness. Following out this line of thought, Sellars takes it that "grasp of a concept is mastery of the use of a word." He then understands the metalinguistic features in question in terms of rules of *inference,* whose paradigms are Carnap's L-rules and P-rules. His generic term for the inferences that articulate the contents of ordinary empirical descriptive concepts is "material inferences." The term is chosen to contrast with inferences that are 'formal' in the sense of depending on *logical* form. In another early essay he lays out the options he considers like this:

> . . . we have been led to distinguish the following six conceptions of the status of material rules of inference:
>
> (1) Material rules are as essential to meaning (and hence to language and thought) as formal rules, contributing to the architectural detail of its structure within the flying buttresses of logical form.
> (2) While not essential to meaning, material rules of inference have an original authority not derived from formal rules, and play an indispensable role in our thinking on matters of fact.
> (3) Same as (2) save that the acknowledgment of material rules of inference is held to be a dispensable feature of thought, at best a matter of convenience.
> (4) Material rules of inference have a purely derivative authority, though they are genuinely rules of inference.
> (5) The sentences which raise these puzzles about material rules of inference are merely abridged formulations of logically valid inferences. (Clearly the distinction between an inference and the formulation of an inference would have to be explored).
> (6) Trains of thought which are said to be governed by "material rules of inference" are actually not inferences at all, but rather activated associations which mimic inference, concealing their intellectual nudity with stolen "therefores."[15]

His own position is that an expression has conceptual content conferred on it by being caught up in, playing a certain role in, material inferences:

> . . . it is the first (or "rationalistic") alternative to which we are committed. According to it, material transformation rules determine the descriptive meaning of the expressions of a language within the framework provided by its logical transformation rules. . . . In traditional language, the "content" of concepts as well as their logical "form" is determined by the rules of the Understanding.[16]

By "traditional language" here, he means Kantian language. The talk of "transformation rules" is, of course, Carnapian. In fact in this essay Sellars identifies his "material rules of inference" with Carnap's "P-rules." 'Determine' is crucially ambiguous between 'constrain' and 'settle'—the difference corresponding to that between what I have elsewhere called 'weak' and 'strong' semantic inferentialism.

As already indicated, the material inferential rules that in one or another of these senses "determine the descriptive meaning of expressions" are for Sellars just the subjunctively robust, hence explanation-supporting ones. As he puts the point in the title of a long essay, he construes "Concepts as Involving Laws, and Inconceivable without Them." This is his response to Quine's implicit challenge in "Two Dogmas of Empiricism" to say what feature of their use distinguishes inferences determining conceptual contents from those that simply register matters of fact. Since empirical inquiry is generally required to determine what laws govern concepts such as *copper, temperature,* and *mass,* Sellars accepts the consequence that it plays the role not only of determining facts but also of improving our conceptions—of teaching us more about the concepts that articulate those facts by teaching us more about what really follows from what.

On this way of understanding conceptual content, the modal concepts that express the lawfulness of connections among concepts and so underwrite subjunctively robust implications—concepts such as *law, necessity,* and what is expressed by the use of the subjunctive mood—have a different status from those of ordinary empirical descriptive concepts. Rather than in the first instance describing how the world is, they make explicit features of the framework that makes such description possible. Because they play this distinctive framework-explicating role, what they express must be implicitly understood by anyone who can deploy *any* ground-level descriptive concepts. As I would like to put the point, in knowing how to (being able to) use any ordinary empirical descriptive vocabulary, each interlocutor already knows how to do everything she needs to know how to do in order to be able to deploy the modal locutions that register the subjunctive robustness of the inferences that determine the content of the descriptive concepts that vocabulary expresses. This is what Kant's idea that the pure concepts of the understanding are knowable *a priori* becomes when transposed into Sellars' framework.

The two lines of thought that orient Sellars' treatment of alethic modality, semantic inferentialism and a metalinguistic understanding of the expressive role characteristic of modal locutions, are epitomized in an early formulation:

> I shall be interpreting our judgments to the effect that A causally necessitates B as the expression of a rule governing our use of the terms 'A' and 'B',[17]

where the rule in question is understood as a rule licensing subjunctively robust inferences. I have been filling in the claim that this overall approach to modality deserves to count as a development of Kant's notion of categories, pure concepts of the understanding, as concepts that make explicit features of the discursive framework that makes empirical description possible. Sellars himself, however, does not discuss this aspect of his work under that heading. When he talks about categories he turns instead to his nominalism about abstract entities. The central text here is "Toward a Theory of the Categories" of 1970.[18] The story he tells there begins with Aristotle's notion of categories (though he waves his hands wistfully at a discussion of its origins in Plato's *Sophist* that he feels he cannot shoehorn into the paper) as ontological *summa genera*. There he opposes an unobjectionable hierarchy:

> Fido is a dachshund.
> Fido is a dog.
> Fido is a brute.
> Fido is an animal.
> Fido is a corporeal substance.
> Fido is a substance.

to a potentially problematic one:

> X is a red.
> X is a color.
> X is a perceptual quality.
> X is a quality.[19]

The next decisive move in understanding the latter hierarchy he attributes to Ockham, whom he reads as transposing the discussion into a metalinguistic key. Ockham's strategy, he tells us, is to understand

(A) Man is a species.

as

(B) •Man• is a sortal mental term.[20]

while construing mental items as "analogous to linguistic expressions in overt speech."

This sketch sets up the transition to what Sellars makes of Kant's understanding of categories:

> What all this amounts to is that to apply Ockham's strategy to the theory of categories is to construe categories as classifications of conceptual

> items. This becomes, in Kant's hands, the idea that categories are the most generic functional classifications of the elements of judgments.[21]

At the end of this development from Aristotle through Ockham to Kant, he concludes:

> [I]nstead of being *summa genera* of entities which are objects 'in the world,' . . . categories are *summa genera* of conceptual items.[22]

The account he goes on to expound in this essay, as well as in his other expositions of his nominalism about terms for qualities or properties, construes such terms metalinguistically, as referring to the inferential roles of the base-level concepts as used in empirical descriptions. I explain how I understand the view and the arguments on this topic in "Sellars's Metalinguistic Expressivist Nominalism." Without going into that intricate view further here, the point I want to make is that although Sellars does not say so, the metaconceptual role he here explicitly puts forward as a successor-concept to Kant's notion of *category* is generically the same as that I have argued he takes alethic modal locutions to play. It is this capacious conception I want to build upon and develop further.

4. Categories Today

The general conception of pure categorial concepts that I have been attributing to Sellars, based on the commonalities visible in his treatment of alethic modal vocabulary and of abstract ontological vocabulary, develops Kant's idea by treating some vocabularies (and the concepts they express) as "covertly metalinguistic." This Sellarsian conception represents his development of Carnap's classification of some expressions as "quasi-syntactic." The underlying insight is that some important kinds of vocabularies that are not strictly or evidently metalinguistic are used not (only) to describe things, but in ways that (also) depend on the use of *other* vocabularies—paradigmatically, empirical descriptive ones.

The lessons I draw from the strengths and weaknesses of Sellars' successor-conception of the "pure concepts of the Understanding" are fourfold. That is, I think he is pointing toward an expressive role characteristic of some concepts and the vocabularies expressing them that has four distinctive features. First, these concepts express what I will call "pragmatically mediated semantic relations" between vocabularies. Second, these concepts play the expressive role of making explicit essential features of the use of some other vocabulary. Third, the proper use of these concepts can be systematically elaborated from the use of that other vocabulary. Fourth, the features of vocabulary (concept)-use they explicate are universal: they are features of any and every autonomous discursive practice. I think there are concepts that play this distinctive fourfold expressive role,

and that a good thing to mean today by the term "category" is metaconcepts that do so.

Carnap and Tarski introduced the expression "metalanguage" for languages that let one talk about languages, with the examples of syntactic and semantic metalanguages. In his earliest writings, Sellars also talks about "pragmatic metalanguages," meaning languages for talking about the *use* of languages—rather than the syntactic or semantic properties of expressions.

These were to be the languages in which we conduct what he called "pure pragmatics." During and after Sellars' most important work in the *anni mirabiles* of 1954–63, however (possibly influenced by Carnap), he shifts to using the expression "semantics" to cover essentially the same ground. I think that this was a step backward, and that it is one of the obstacles that prevented him from getting clear about the sense in which he wanted to claim that such locutions as alethic modal vocabulary and singular terms purporting to refer to universals ("circularity") and their kinds ("property") are "covertly metalinguistic." One vocabulary serving as a pragmatic metavocabulary for another is the most basic kind of pragmatically mediated semantic relation between vocabularies. It deserves to be called such because the *semantics* of the pragmatic metavocabulary depends on the *use* of the vocabulary for which it is a pragmatic metavocabulary. The relation itself is aptly called a "semantic" relation in the special case where one vocabulary is sufficient to specify practices or abilities whose exercise is sufficient to confer on another vocabulary the meanings that it expresses. We could represent such a semantic relation, mediated by the practices of using the second vocabulary that the first vocabulary specifies, as in Figure 12.1.[23]

The pragmatically mediated semantic relation between vocabularies V′ and V, indicated by the dashed arrow, obtains when vocabulary V′ is expressively sufficient to *specify* practices-or-abilities P (that semantic fact about V′ with respect to P is here called "VP-sufficiency") that are sufficient to

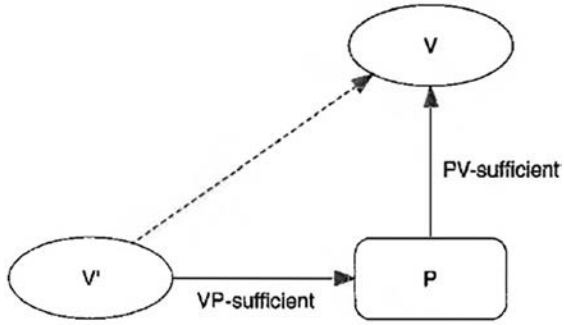

Figure 12.1 Meaning-use diagram representing the pragmatically mediated semantic relation of a pragmatic metavocabulary, V′, to another vocabulary, V. P are the practices-or-abilities to deploy the vocabulary V.

deploy the vocabulary V with the meanings that it expresses when so used. In asserting that this relation between vocabularies obtains, one is claiming that if all the sentences in V′ used to specify the practices-or-abilities P are true of P, then anyone engaging in those practices or exercising those abilities as specified in V′ is using the expressions of V with their proper meanings. This semantic relation between what is expressible in the two vocabularies is mediated by the practices P that the first specifies and that are the use of the second. This particular pragmatically mediated semantic relation holds when the vocabulary V′ allows one to *say* what one must *do* in order to *say* what can be said in the vocabulary V. In that sense V′ makes *explicit* (sayable, claimable) the practices-or-abilities *implicit* in using V. This is the explicative relation I mention as the second component of the complex expressive role that I am offering as a candidate for a contemporary successor-(meta)concept to Kant's (meta)concept of *category*. There are other pragmatically mediated semantic relations besides being a pragmatic metavocabulary in this sense, and others are involved in the categorial expressive role. The result will still fall under the general rubric that is the first condition: being a pragmatically mediated semantic relation.

One such further pragmatically mediated semantic relation between vocabularies holds when the practices PV-sufficient for deploying one vocabulary, though not themselves PV-sufficient for deploying a second one, can be systematically elaborated into such practices. That is, in being able to deploy the first vocabulary, one already knows how to do everything one needs to know how to do, in principle, to deploy the second. But those abilities must be suitably recruited and recombined. The paradigm here is *algorithmic* elaboration of one set of abilities into another. Thus, in the sense I am after, the capacities to do multiplication and subtraction are algorithmically elaborable into the capacity to do long division. *All* you need to learn how to do is to put together what you already know how to do in the right way—a way that can be specified by an algorithm. The diagram for this sort of pragmatically mediated semantic relation between vocabularies is shown in Figure 12.2.

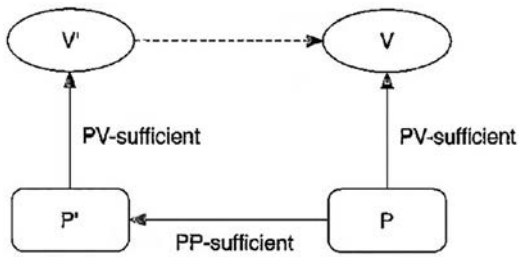

Figure 12.2 Meaning-use diagram representing the relation between two vocabularies, V′ and V, that holds if the practices-or-abilities sufficient to deploy vocabulary V can be elaborated into practices sufficient to deploy vocabulary V′.

The dotted arrow indicates the semantic relation between vocabularies V′ and V. It is the relation that holds when all the relations indicated by solid arrows hold—that is, when the practices-or-abilities sufficient to deploy vocabulary V can be elaborated into practices sufficient to deploy vocabulary V′. In this case, the semantic relation in question is mediated by two sets of practices-or-abilities: those sufficient to deploy the two vocabularies.

A concrete example of vocabularies standing in this pragmatically mediated semantic relation, I claim, is that of *conditionals* in relation to ordinary empirical descriptive (OED) vocabulary. For using such OED vocabulary, I claim (following Sellars following Kant), requires distinguishing in practice between materially good inferences involving descriptive predicates and ones that are not materially good. One need not be either infallible or omniscient in this regard, but unless one makes *some* such distinction, one cannot count as deploying the OED vocabulary in question. But in being able practically to distinguish (however fallibly and incompletely) between materially good and materially bad inferences, one knows how to do everything one needs to know how to do, in principle, to deploy conditionals. For conditionals can be introduced by recruiting those abilities in connection with the use of sentences formed from the old vocabulary by using the new vocabulary. On the side of circumstances of application (assertibility conditions), one must acknowledge commitment to the conditional p→q just in case one takes the inference from p to q to be a materially good one. And on the side of consequences of application, if one acknowledges commitment to the conditional p→q, then one must take the inference from p to q to be a materially good one. These rules constitute an algorithm for elaborating the ability to distinguish materially good from materially bad inferences using OED vocabulary (or any other vocabulary, for that matter) into the ability appropriately to use conditionals formed from that vocabulary: to distinguish when such conditionals are assertible, and what the consequences of their assertibility is.

My idea for a successor-concept to what Sellars (with hints from Carnap) made of Kant's metaconception of pure concepts of the Understanding is that they must play *both* of these expressive roles, stand in *both* sorts of pragmatically mediated semantic relations to another vocabulary. It must be possible to *elaborate* their use from the use of the index vocabulary, and they must *explicate* the use of that index vocabulary. Speaking more loosely, we can say that such concepts are both *elaborated from* and *explicative of* the use of other concepts—in short that they are el-ex, or just LX with respect to the index vocabulary.

The fourth condition I imposed above is that the concepts in question must be *universally* LX, by which I mean that they must be LX for every autonomous discursive practice (ADP)—every language game one could play though one played no other. That is, the practices from which their use can be elaborated and of which their use is explicative must be essential

to talking or thinking at all. This universality would distinguish categorial concepts, in the sense being specified, from metaconcepts that were elaborated from and explicative of only some parasitic fragment of discourse—culinary, nautical, or theological vocabulary, for instance. I take it that any autonomous discursive practice must include the use of ordinary empirical descriptive vocabulary. If so, being LX for OED vocabulary would suffice for being *universally* LX, LX for every ADP.

Putting all these conditions together yields the diagram (shown in Figure 12.3) of the pragmatically mediated semantic relation between vocabularies that obtains when vocabulary V′ plays the expressive role of being universally LX by being elaboratable from and explicative of practices necessary for the deployment of ordinary empirical descriptive vocabulary.

The fact that the rounded rectangle labeled P″, representing the practices from which vocabulary V′ is elaborated and of which it is explicative, appears inside the rounded rectangle representing practices sufficient to deploy ordinary empirical descriptive vocabulary indicates that the practices P″ are a necessary part of the practices sufficient to deploy OED vocabulary, but need not comprise all such practices. Thus, distinguishing materially good from materially bad inferences involving them is necessary for deploying ordinary empirical descriptive vocabulary (rather than mere labels), but there is a lot more involved in doing so—using such vocabulary observationally, for instance. Different categorial metaconcepts can be LX for different essential features of the use of empirical descriptive vocabulary. Thus alethic modal vocabulary explicates the subjunctive robustness of the inferences explicated by conditionals. "Quasi-syntactic" abstract ontological vocabulary such as 'property' and 'proposition' explicates structural features of descriptive sentences.

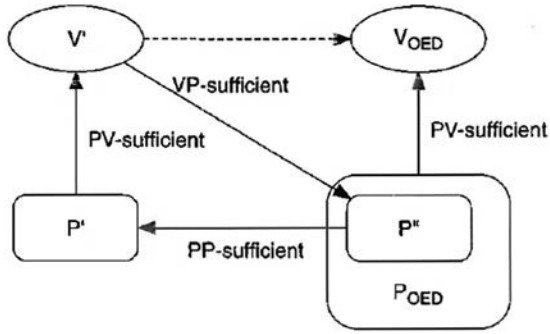

Figure 12.3 Meaning-use diagram representing the pragmatically mediated semantic relation between vocabularies that obtains when vocabulary V′ plays the expressive role of being universally LX.

Diagramming the expressive role of being LX for practices necessary to deploy OED vocabulary provides an analysis that breaks down the claim that some vocabulary plays a categorial role into its component sub-claims. To show that alethic modal vocabulary, for instance, stands in this pragmatically mediated semantic relation to ordinary empirical descriptive vocabulary one must show that there are some practices-or-abilities (in this case, to reason subjunctively or counterfactually) that are 1) a necessary component of practices-or-abilities that are 2) PV-sufficient to deploy OED vocabulary, 3) from which one can elaborate practices-or-abilities that are 4) PV-sufficient to deploy vocabulary (alethic modal vocabulary) 5) that is VP-sufficient to explicate or specify the original practices-or-abilities. Although there is by design considerable elasticity in the concepts *vocabulary, practices-or-abilities,* and the various sufficiency and necessity relations between them, the fine structure of the distinctive expressive role in question is clearly specified.

What credentials does that expressive role have to pick out a worthy successor metaconcept to what Sellars made of Kant's *categories* or *pure concepts of the Understanding?* At the beginning of my story, I introduced the idea behind the Kantian categories as the idea that besides the concepts whose principal use is in giving empirical descriptions and explanations, there are concepts whose principal use is in making explicit features of the framework that makes empirical description and explanation possible. The expressive task characteristic of concepts of this latter class is to articulate what Kant called the "transcendental conditions of experience." The concepts expressed by vocabularies that are LX for empirical descriptive vocabulary perform this defining task of concepts that are categories. As explicative of practices necessary for deploying vocabularies performing the complex expressive task of description and explanation (distinguishable only in the context of their complementary relations within a pragmatic and semantic context that necessarily involves both), this kind of vocabulary makes it possible to *say* what practitioners must be able to *do* in order to describe and explain how things empirically are. They do this by providing a pragmatic metavocabulary for describing and explaining. This is a central feature (the 'X' in 'LX') of the complex pragmatically mediated semantic relation between categorial metaconcepts and ordinary empirical descriptive vocabulary.

One feature of the concepts performing this explicative function that Kant emphasizes is that they are "*pure* concepts of the Understanding." (I take it that the "of" should be understood as expressing both the subjective and objective genitives—as in "Critique of Pure Reason." These concepts both belong to the Understanding and address it, being both discursive and metaconceptual.) To say that they are pure concepts is to say that they are graspable *a priori*.[24] The feature of the LX model that corresponds to the *a prioricity* of Kant's categories is that the use of LX metaconcepts can be elaborated from that of the empirical descriptive vocabularies for which

they are LX. As I have put the point, in knowing how to deploy OED vocabulary, one already knows how to do everything one needs to know how to do to deploy vocabulary that is LX for it—such as alethic modal vocabulary, conditionals, and ontological classificatory vocabulary. If we take it, as per Sellars, that grasp of a concept is mastery of the use of a word, then one need not *actually* grasp concepts that are LX for descriptive vocabulary in order to deploy descriptive vocabulary. But in effect, *all* one is missing are the words for them. The circumstances and consequences of application of LX concepts can be formulated by rules that appeal only to abilities one already has in virtue of being able to use OED vocabulary. (Think of the sample rules for conditionals sketched above.) In that sense, the LX concepts are *implicit in* the descriptive concepts. It is not that one must or could grasp these concepts *before* deploying descriptive concepts. It is rather that nothing more is required to grasp them than is required to deploy descriptive concepts, and there are no particular descriptive concepts one must be able to deploy, nor any particular descriptive claims that one must endorse, in order to possess abilities sufficient to deploy the universally LX metaconcepts.

The class of concepts that are arguably universally LX (LX for every autonomous discursive practice because LX for OED vocabulary) overlaps Kant's categories in important ways—most notably in the alethic modal concepts that make explicit subjunctively robust consequential relations among descriptive concepts. But the two do not simply coincide. Besides modal vocabulary, as I argue in *Between Saying and Doing*, logical vocabulary, indexical and demonstrative vocabulary, normative vocabulary, and semantic and intentional vocabulary all should be thought of as LX for OED vocabulary. In spite of this extensional divergence, the fact that vocabulary that is LX for descriptive vocabulary in general principle shares with Kant's categories the two crucial features of being explicative of such vocabulary and being graspable *a priori* makes the idea of universally LX metaconcepts a worthy successor to Kant's breakthrough idea. The fact that Sellars' own development of this idea of Kant's takes such important steps in this direction convinces me that his version of the categories was a progressive step, and a Good Idea.

Notes

1 In Herbert Feigl and Michael Scriven (eds.), *Minnesota Studies in the Philosophy of Science,* vol. I (Minneapolis: University of Minnesota Press, 1956); reprinted in Sellars's *Science, Perception, and Reality* (London: Routledge and Kegan Paul, 1956; reissued Atascadero, CA: Ridgeview, 1991); reprinted as a monograph, with an Introduction by Richard Rorty and a Study Guide by Robert Brandom (Cambridge, MA: Harvard University Press, 1997). Hereafter *EPM*.

2 In his introduction to my Harvard University Press edition of *Empiricism and the Philosophy of Mind*.

3 His only rival for this accolade, I think, would be Peter Strawson, who certainly did a lot to make us realize that a reappropriation of some of Kant's theoretical philosophy might be a viable contemporary project. But I do not think of Peter Strawson's work as *systematically* neo-Kantian in the way I want to argue that Sellars' is.

4 Paul Redding begins the process of recovering the necessary counter-narrative in the Introduction to his *Analytic Philosophy and the Return of Hegelian Thought* (Cambridge: Cambridge University Press, 2010).

5 Distinguishing two broadly different kinds of *use* bits of vocabulary can play does not entail that there are two corresponding kinds of *concepts*—even in the presence of the auxiliary Sellarsian hypothesis that grasp of a concept is mastery of the use of a word. Though I suppress the distinction between these two moves in these introductory formulations, it will become important later in the story.

6 "Counterfactuals, Dispositions, and the Causal Modalities," in H. Feigl, M. Scriven, and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. II (Minneapolis: University of Minnesota Press, 1957), §79. Hereafter CDCM.

7 The best candidate might be the discussion of "hinge propositions" in *On Certainty*. But the point there is, I think, different. In any case, Wittgenstein does not *generalize* the particular expressive role he is considering to anything like the extent I am claiming Sellars does.

8 Note that these concepts are *not* those Kant discusses under the heading of "Modality" but rather concern the hypothetical form of judgment.

9 I take it that Kant always uses "*a priori*" and "*a posteriori*" as adverbs, modifying some verb of cognition, paradigmatically "know."

10 In Chapter 3 of *From Empiricism to Expressivism: Brandom Reads Sellars*, I discuss the sense in which "metalinguistic" should be understood in such formulations. See R. B. Brandom, *From Empiricism to Expressivism: Brandom Reads Sellars* (Cambridge, MA: Harvard University Press, 2015).

11 Chapter 7 of Brandom's *From Empiricism to Expressivism* discusses Sellars' view about this kind of locution.

12 "Another feature of the empiricist tradition is its 'logical atomism,' according to which every basic piece of empirical knowledge is logically independent of every other. Notice that this independence concerns not only *what* is known, but the *knowing* of it. The second dimension of this 'atomism' is of particular importance for understanding Kant's rejection of empiricism. . . ." Sellars, "Toward a Theory of the Categories," in *Essays in Philosophy and Its History* (Dordrecht: D. Reidel, 1974), §16.

13 CDCM §108.

14 R. Carnap, *The Logical Syntax of Language* (London: Kegan Paul, 1937), §§63–70.

15 Wilfrid Sellars, "Inference and Meaning," in J. Sicha (ed.), *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars* (Atascadero, CA: Ridgeview, 1980), Reprinted in Kevin Scharp and Robert Brandom (eds.), *In the Space of Reasons: Selected Essays of Wilfrid Sellars* (Cambridge, MA: Harvard University Press, 2007), p. 317. Hereafter *PPPW*.

16 Sellars, "Inference and Meaning," *PPPW*, p. 336.

17 Sellars, "Language, Rules, and Behavior," *PPPW*, footnote 2 to p. 296.

18 In L. Foster and J. W. Swanson (eds.), *Experience and Theory* (Amherst: University of Massachusetts Press, 1970), pp. 55–78; reprinted in *Essays in Philosophy and Its History*. Hereafter TTC.

19 TTC §10–11.

20 TTC §16.

21 TTC §22.
22 TTC §23.
23 I introduce, develop, and apply these "meaning-use diagrams" in *Between Saying and Doing: Towards an Analytic Pragmatism* (Oxford: Oxford University Press, 2008).
24 Kant does admit also impure *a priori* principles.

Part IV

Author Meets Critics

Robert B. Brandom, Willem A. deVries, and James O'Shea

The following are the initial remarks, responses, and subsequent dialogue among Robert B. Brandom, Willem A. deVries, James O'Shea, and other participants on Robert B. Brandom's *From Empiricism to Expressivism: Brandom Reads Sellars* from their May 2, 2015, presentation at the Sellars in a New Generation conference at Kent State University.

BOB BRANDOM: I want to say something in general about why I wrote this book. I talked about two neo-Kantian ideas—the White Hat and the Black Hat—already (in chapter 12 of this volume). I've been worrying about Sellars my entire career, and there is stuff I had given up on understanding. I think I'm together on the early Sellars, the middle Sellars of this fabulous nine years starting in 1956, of the late Sellars when he was my colleague, and the meshugganah Sellars of pure processes as well. But absolutely at the core of his system, I came to think, was his nominalism about abstract entities, and I could *never* understand why he thought this was so important. Modality: I could see why that was really important—both in the sociology of philosophy, but downstream from Kant, you have to get straight on modality. But all we have on modality [in Sellars] are some scattered references, mostly in the early work, and then "Counterfactuals, Dispositions and the Causal Modalities," which is a mess. I don't know why the first half of that essay exists, and the second half breaks off without him having resolved the problems he is settling. He never returns to it; I think he didn't know what to say about modality. By contrast, we have the three big essays on nominalism. One of them, "Grammar and Existence," is basically two long essays, so there are really the four of them. In a three-year period he basically wrote a book on this topic. He really did nail down what he needed to say, and his later references to this work indicate he was completely happy with the view. This stands in contrast to his treatment of modality. I couldn't see why he cared about it, why he thought this was so central outside of the epistemological and semantic moves in EPM, which he continued to refer to; nominalism, Sellars clearly thought, was at the center of what he was doing.

It was only when I was able to see this as a principled application of this metalinguistic expressive strategy for dealing with philosophically puzzling concepts that I said, "All right—this one he really worked out!" Sellars showed how this metalinguistic strategy could work: Thinking about these statements as *what you're doing when you say something is a quality or property* is classifying the expression that is being nominalized as a predicate, not as a singular term. So I wanted to focus the book on this metalinguistic expressive strategy.

I think that Sellars was entirely successful in this nominalism. He sets out criteria of adequacy, and he satisfies them. But there is a dark side to this, too. Namely, he thought that when you give an expressivist account of something that means that it doesn't play any descriptive or representational role. So, if as he thought—and as I think—he gave us a correct expressivist account of concepts like *particular*, *object*, *property,* and so on, but in particular of *concepts that name properties* (like 'circularity', and so on), then his conclusion was that those things are not real in the narrow sense. They don't exist in the world, in the narrow sense. There is a broader sense in which they do, but that is the broader sense in which norms exist in the world. But they're not describable, representable—they are not in the world *in the narrow sense*. This bleeds into Sellars' discussion of scientific naturalism, because that world in the narrow sense—what's real, what exists in the narrow sense—that is what the language of natural science is supposed to be authoritative about. But it is not a world of facts; it is a world of things. It is just the particulars. There are no properties in it, there are no facts in it. It is just the things. Wittgenstein said that there is an opposition in thinking of the world as a world of things, and a world of facts, and comes down on the facts side. This view is completely explicit in "Abstract Entities," the last of his [Sellars'] nominalism essays. It is actually reflected in the *scientia mensura*. Now this has come up many times; I'll bet you're sick of hearing about it! I'd like to ask a question that I'll bet you've never asked yourself before: Why isn't the *scientia mensura* the following: "In the dimension of describing and explaining, science is the measure of all things, of what is, and what it is, and of what is not, what it is not"? All he says is "of what is" and "that it is not." Now, right off the bat, there are two readings of that. One is to say that it is just stylistic, he's doing a pastiche of Protagoras, who spoke of what is and what is not, and Sellars, for stylistic reasons, didn't want to break that. In which case, the missing piece of the *scientia mensura* is of no significance at all. But it actually expresses this nominalistic view that the world, in the narrow sense—what's real, what exists in the narrow sense—consists exclusively of particulars. Late in life, he comes to think that those particulars are pure processes, but he doesn't give up the nominalism; we still don't have properties of them in the world, we still don't have facts, we just have things. I doubt that this conception of the world as just a

collection of things is intelligible, but I am confident that the implicit argument that got him there is a bad argument. The argument is that he notices—and this is a wonderful thing to notice—that these expressions, these ontological categorizing expressions—like 'circularity', like 'property'—play this distinctive expressive role that is not played by ordinary empirical descriptive vocabulary. There is a distinction in their expressive role. He concludes that they do not play a descriptive role, not even in a broader sense of descriptive. They do not represent how things are. Now that made sense if you thought that the sense in which they were metalinguistic was a semantic metalanguage that they were expressed in. But that didn't quite work for him. That's why he had to say that you're not *saying* that the inference is good, or that this expression is a predicate. You're *conveying* that information. I've suggested that his better wisdom—what he is reaching for there—is what he has a fabulous account of: In a pragmatic metalanguage, rather than a semantic one, he describes what you're *doing* in saying what something means. You're functionally classifying it, or in saying what ontological category it is, you're functionally classifying it in a quasi-syntactic way. The inference from what you're doing when you say *this* is something you don't do when you use ordinary empirical descriptive vocabulary. So you can't be saying anything of the kind that ordinary empirical descriptive vocabulary does. You can't be representing the way the world is. That's just a bad inference. It doesn't follow from *saying what you're doing* when you say "something is this"—nothing actually follows about *what you're saying* when you say that. And yet that's the inference: He says what you're not doing is saying anything about what's real, you're not representing the way the world really is.

It is clear that this nominalism that is at the center of Sellars' thought is, in some sense, the best worked out philosophy that he did. But it led him to this conclusion that what exists in the narrow sense, that what is real, is just a collection of particulars. This is a very popular view. My other "Doktorvater," David Lewis, says what's real is whatever the ultimate particles are, and all the mereological sums of them. That's thinking of the world as just a collection of particulars. The push-back you get to that view in contemporary metaphysics is, "Oh, no, it is world just of things but it's gunk, there aren't ultimate particles," *not* [the view] that the only conception of a world that makes sense is a world of facts, of propertied particulars. I myself doubt that this reistic conception—often associated with Kotarbiński, whom Sellars sometimes talks about—I doubt that this [view] is intelligible. But I am much more confident that his inference from his expressivist analysis that I understand as a matter of *what you're doing* when you say something—that the inference from that to "so you're not doing anything that's descriptive" (even in the broader sense than the narrow sense that applies to ordinary empirical descriptive vocabulary)—I'm confident that is not a good inference.

When we think about his scientific naturalism, the *first step he makes* is thinking of the world as the world of particulars first, and then that's what science is authoritative about. I think the second move is tainted by the first move. At any rate, if we're going to get on to figuring out what's living and what's dead in Sellars, that's something we have to worry about: In what sense is the world a world of facts, and in what sense is it a world of particulars?

Now Bill and Jim are going to talk about the scientific naturalism, not this particular aspect of it, but I didn't want to let us not have that on the table.

BILL DEVRIES: Both Jim and I both thought we had to say *something* in favor of the Black Hat, because in the long run it may be the White Hat that wins and survives, but the Black Hat should at least go down fighting, right?

The last half of the (long) first chapter of Brandom's *From Empiricism to Expressivism* constitutes an extended argument against one half of Wilfrid Sellars' version of scientific realism. I say 'half' of Sellarsian scientific realism because Brandom agrees with Sellars' anti-instrumentalism. The half Brandom takes issue with is Sellars' claim that the "scientific image" [SI]—an idealized, complete scientific framework for the description and explanation of all natural events and objects—possesses such ontological priority over the "manifest image" [MI]—itself an idealization of the 'commonsense' framework of persons and things in terms of which we currently experience ourselves and the world—that it will come to replace the MI in all matters of description and explanation.

Brandom's argument against this Sellarsian idea is rather roundabout. First, he traces Sellars' distinction between the MI and the SI back to the Kantian distinction between phenomena and noumena. Then he argues against several attempts to understand identity claims across disparate frameworks. Neither, claims Brandom, will permit us to identify objects across the MI/SI divide. But if we cannot identify the objects of concern across the frameworks, well then a shift from the MI to the SI is not a form of *replacement* of one framework by a better, but simply a change of subject, and that poses no threat to the MI at all. And thus Brandom strikes a blow for left-wing Sellarsianism.

Now I'm not going to argue that Brandom is wrong to thus defend left-wing Sellarsianism. Nor am I going to argue that Brandom is wrong to see Sellars' MI/SI distinction as a latter-day version of Kant's phenomena/noumena distinction. I too am bothered by Sellars' belief that the scientific image will be able to replace the manifest image in some wholesale fashion (some of the difficulties of this view is the subject of my paper in this volume). But I do think that Brandom has *not yet* sealed the deal. There is another construal of Sellars' distinction that I think could survive Brandom's arguments, and it is not a construal that

is obvious on the face of it; nor, for that matter, can I really claim that it is the "real Sellarsian construal"—the one that he had in mind. But I do think it is a construal of Sellars' doctrine that is intrinsically interesting, and it is a construal that, given Brandom's endorsement of Sellars on categories, and given, further, Brandom's own distinctively Hegelian leanings, Brandom needs to face.

(I wrote this when I first got the book, so I have to skip all the pages where I pat Brandom on the back for doing all the stuff about the categorial Sellars that he likes—'cause I like that too—but this is a *critics* session, right?)

Sellars introduces the manifest image in existential terms, as "the framework in terms of which man came to be aware of himself as man-in-the-world" (PSIM ¶14; SPR: 6; in ISR: 374). Though he contrasts it with the scientific image, the manifest image is neither uncritical nor naïve and unscientific. It has been refined over the millennia both categorially and empirically. In thinking about the nature of a conceptual framework, Sellars thinks one of the fundamental questions to ask is "'of what sort are the basic objects of the framework?'" (PSIM ¶26; in SPR: 9; in ISR: 377). For the manifest image, the answer is persons and things. It's a Strawsonian, or Aristotelean-Strawsonian, framework. There is a fairly complex (and not altogether plausible) backstory attached to this claim, wherein we began with an 'original image' (and I've never quite got how that could be the original one) in which everything is accounted a way of being a person. The notion of a thing develops as we come to realize that not everything exhibits the full range of capacities that characterize persons.

Thus, the manifest image includes what Sellars calls the "descriptive ontology of everyday life" (EPM §41). "Perennial philosophy," he tells us, "which is the 'ideal type' around which philosophies in what might be called, in a suitably broad sense, the Platonic tradition cluster, is simply the manifest image endorsed as real, and its outline taken to be the large-scale map of reality to which science brings a needle-point of detail and an elaborate technique of map-reading" (PSIM ¶21: in SPR: 8; in ISR: 376). That's how he takes perennial philosophy to see things. The differences between the SI and the MI are generated from a single, methodological difference: The scientific image begins to form when we begin to postulate imperceptible entities to explain the behavior of perceptible things. This entails, of course, that the scientific image presupposes the prior availability of the manifest image in terms of which we perceive things in the first place, but it is crucial to Sellars' view that the methodological priority of the manifest image does not imply its substantive or ontological priority: what is first in the order of knowing need not be first in the order of being.

The overall story Sellars tells is then fairly clear: in the process of postulating imperceptible entities to explain the observable behavior of

things, we do not simply add more of the same kinds of things already believed in to our world-view; we add *new* kinds of things and sometimes, correlatively, new kinds of concepts. Science revises the categorial structure of our world-view. Sellars sees this, ultimately, as a challenge to the manifest image. We cannot simply add new categories to our framework; we must also prune the old. Sellars is radical here; he does not think we can prune and replace in a piecemeal fashion: "[T]he most fruitful way of approaching the problem of integrating theoretical science with the framework of sophisticated common sense into one comprehensive synoptic vision is to view it not as a piecemeal task—e.g. first a fitting together of the common sense conception of physical objects with that of theoretical physics, and then, as a separate venture, a fitting together of the common sense conception of man with that of theoretical psychology—but rather as a matter of articulating two whole ways of seeing the sum of things, two images of man-in-the-world and attempting to bring them together in a 'stereoscopic' view" (PSIM, ¶52; in SPR: 19; in ISR: 387).

Notice that the assumed unity of science plays a significant role here. Sellars is not totally naïve on that score; he's not totally naïve about much. He certainly recognizes the methodological pluralism of the sciences; he sees that "as sciences they have different procedures and connect their theoretical entities via different instruments to intersubjectively accessible features of the manifest world" (PSIM ¶58; in SPR: 21; in ISR: 389). "But" Sellars argues, "diversity of this kind is compatible with intrinsic 'identity' of the theoretical entities themselves, that is, with saying that biochemical compounds are 'identical' with patterns of subatomic particles. For to make this 'identification' [and 'identification' is in quotes there, for some reason] is simply to say that the two theoretical structures, each with its own connection to the perceptible world, could be replaced by one theoretical framework connected at two levels of complexity via different instruments and procedures to the world as perceived" (PSIM ¶58; in SPR: 21; in ISR: 389). Identity claims will play, obviously, a significant role in the arguments to come.

Sellars spells out more thoroughly what this means, in his view, in *Science and Metaphysics*: "A consistent scientific realist must hold that the world of everyday experience is a phenomenal world in the Kantian sense, existing only as the contents of actual and obtainable conceptual representings, the obtainability of which is explained not, as for Kant, by things in themselves known only to God, but by scientific objects about which, barring catastrophe, we shall know more and more as the years go by" (SM VI ¶61: 173). Insofar as Sellarsian scientific realism goes beyond a rejection of scientific instrumentalism (which of course Brandom also rejects), Brandom takes Sellars' position to be, in the end, a reductive scientific naturalism, because science retains an absolute priority in ontology.

So Brandom's criticism of Sellars' scientific realism is intended to be a rejection of reductive scientific naturalism, and I have no interest in defending such a reductionism. It isn't clear to me, however, that Sellars ever espoused that position, and some of Brandom's characterizations of the scientific image that support such an interpretation seem clearly off the mark. So let me pause to correct some mistaken assertions Brandom makes in his characterization of the scientific image.

My primary complaint is that according to Brandom, the scientific image "consists exclusively of descriptions and explanations" (Brandom 57), and "[n]ormative vocabulary accordingly is not drawn upon in articulating the scientific image of things. It belongs exclusively to the manifest image" (Brandom 58). I admit that Sellars heavily emphasizes the descriptive and explanatory dimension of science, and that it is no accident that the *scientia mensura* begins with the condition "In the dimension of describing and explaining the world. . . ." But if the SI consists of *nothing* but descriptive/explanatory discourse, I'm not sure why Sellars would need the introductory condition in the *scientia mensura* in the first place, since Sellars claims that the scientific image purports to be a complete image of man-in-the-world. In Brandom's view, however, the scientific image is not only purely descriptive but also shorn of all prescriptive discourse. It would be deeply incoherent for Sellars, who is so very sensitive to the rich multidimensionality of language and the conceptual frameworks that define the structures of thought, to think that there could be a "complete" image of the world that contains and employs only descriptive vocabulary. Sellars hints at this when he points out that "the conception of the scientific or postulational image is an idealization in the sense that it is a conception of an integration of a manifold of images, each of which is the application to man of a framework of concepts which have a certain autonomy. For each scientific theory is, from the standpoint of methodology, a structure which is built at a different 'place' and by different procedures within the intersubjectively accessible world of perceptible things" (PSIM ¶55; in SPR: 20; in ISR: 388). So, science(s) clearly, in his characterization, has (have) methodologies and procedures, and those are essentially normative: they tell us what we ought and ought not to do.

I have argued elsewhere that the kind of view of the scientific image that Brandom is trying to sell us, however much it seems supported by Sellars' own words, cannot be right. Sellars talks of the need to "join" the language of individual and community intentions, which provides, in his view, the basis for normative discourse, to the scientific image. Talk of 'joining' implies that the things joined have existence independently of each other, but in my view such talk is really misleading. A scientific image of man-in-the-world—I'm sorry, humanity-in-the-world (sometimes one falls into bad old traditions)—can neither develop nor sustain itself independently of normative language and categories. So I

have to reject Brandom's crass distinction between the two images, and especially his assertion that normative vocabulary "belongs exclusively to the manifest image." Sellars tells us it's got to be added to the SI, and will thereby become part of the scientific image.

There is, however, plenty of reductionist rhetoric in Sellars, and Brandom's attack on this thread in Sellars' thought is welcome as a counterbalance to that rhetoric. The larger question we need to face is whether that rhetoric expresses something deeply woven into the texture of Sellars' thought or is, instead, a more superficial aspect of Sellars' response to the philosophical problems he faced. According to Brandom, Sellars drew inspiration for his MI/SI distinction from Kant's distinction between the phenomenal and the noumenal realms. "The question Sellars' neo-Kantian reappropriation of the phenomena/noumena distinction addresses is how to understand the relations between the descriptive vocabulary native to the manifest image and the descriptive vocabulary native to the scientific image" (Brandom 62). Brandom takes the *scientia mensura* to mean "that descriptive terms from the manifest image refer to things specifiable in descriptive terms from the scientific image, if they refer at all" (Brandom 62). This reading leads to the first construal of Sellars' distinction, which Brandom calls the "sense-reference scientific naturalist rendering of the phenomena/noumena distinction."

Given popular extensionalist assumptions, co-reference of terms is identity of objects, and we've already seen that Sellars is concerned about the identifiability of objects across scientific theories. If we so construe the MI/SI distinction that the relevant descriptive vocabularies constitute two different realms of senses picking out (if anything) a common set of referents, then scientific realism is the view that it is the vocabulary of science that 'really carries all the weight.' It's the only reliable vocabulary we really have. The MI vocabulary successfully refers and the relevant objects exist if and only if the referent is also the referent of (true) scientific assertions. This construal is very congenial to a lot of contemporary scientific naturalists.

Let me try to summarize Brandom's argument very briefly; it is not my purpose here to subject it, as an argument, to close scrutiny. In the standard or received view, identities can be cashed out via Leibniz's law, but with a significant codicil: only extensional predicates count. Extensional predicates are such that what they are true of in a given possible world depends only on what is true in that world. But Brandom argues, "all descriptive properties are modally involved (so that we cannot require that identicals be indiscernible only with respect to modally insulated properties)" (Brandom 76). Furthermore, Brandom claims, "differences in criteria of identity and individuation entail differences in modal profile—that is, differences in the possession of properties whose applicability or possession entails nonmonotonic subjunctive conditionals"

(Brandom 77). This is all stuff you've been hearing—it should be old hat by now. From these premises Brandom concludes "that no identity claims involving terms that fall under descriptive sortals exhibiting different criteria of identity and individuation (that is, no strongly cross-sortal identity claims) are true" (Brandom 76). Brandom recognizes that this is not a knockout argument, because he has not excluded the possibility of constructing some partition of modally involved predicates on which the predicates in one class are referentially transparent and thus some strongly cross-sortal identities come out true. He does think he has made the likelihood of such cases extremely low.

If Brandom's argument is correct, then "the identity version of the sense/reference construal of the scientific naturalist rendering of the phenomena/noumena distinction is untenable, and should be recognized to be so by Sellars' own lights" (Brandom 80). Is there a weaker position then on the distinction between the images that we might attribute to Sellars? Yes, Brandom thinks one in particular stands out: Take functional realization as the basic model of relations between MI and SI. One way to do this is to construe the MI as a theory, Ramsify it (by replacing each bit of descriptive vocabulary in it by a variable bound by a quantifier ranging over predicates or sortals), and then look for the best replacements (realizers) as described in scientific language. Given Sellars' functionalist treatment of the intentional as well as his recognition of the extent to which Aristotle and his hylomorphic metaphysics is a powerful interpretation of the underlying logic of the manifest image, this kind of view slides like a fairly well-fitted glove onto much of what Sellars says. Notice that it also gives us what many have thought to be a reasonable story to tell about the cross-theoretical identities that we saw Sellars worrying about. The identification of genes with DNA sequences, for instance, is ubiquitous, and seems to make sense on that model. But Brandom does not think that this interpretation of the MI/SI relation is available to Sellars, because it conflicts with another compelling Sellarsian argument, namely Sellars' argument against phenomenalism in the essay of that title. Again, I want to be brief in my treatment of Brandom's argument here, because its validity and soundness are not my primary concern. But Brandom tells us, "Both the phenomenalist reductive project and this functionalist rendering of scientific naturalism seek to explain the use of some target vocabulary (object-directed, ordinary empirical description) in terms of the use of a privileged base vocabulary (phenomenal experience talk, scientific description). The phenomenalist looks directly to underwrite subjunctive conditionals whose consequents are expressed in the privileged vocabulary, while the functionalist naturalist looks to reproduce as far as possible the subjunctive conditionals that articulate the criteria of identity and individuation of sortals in the target vocabulary by means of conditionals couched in the privileged vocabulary" (Brandom 83). So Brandom,

following Sellars, poses a dilemma: either all of the target vocabulary is eliminated, or it is not. Suppose we try to Ramsify the manifest image; if we cannot ultimately get rid of it all, the supposed priority of the scientific vocabulary cannot stand. Once one has Ramsified some theory, the general problem, of course, is that there are too many potential realizers or models (e.g., mathematical models). This problem, according to Brandom, is commonly dealt with by requiring that the *causal* relations in the target vocabulary not be Ramsified, which is a reasonable constraint. But then we're pushed onto the other horn of the dilemma: The subjunctive conditionals that the functionalist naturalist seeks to reproduce inevitably include manifest-image sortal vocabulary in their antecedents. Brandom concludes that the functional-realization of the MI/SI relation also fails to hold up. "The result is that the functionalist way of reading Sellars' scientific naturalist rendering of Kant's phenomena/noumena distinction fares no better than the sense/reference identity way of reading it. It just is not the case that everything we talk about in the manifest image that exists at all . . . is something specifiable in the language of an eventual natural science. The manifest image is not best thought of as an appearance of which the world, as described by science, is the reality" (Brandom 87). Brandom's general diagnosis is that Sellars was operating in an atmosphere in which two common assumptions still ruled: (1) a general belief in the unity of science, where that unity is interpreted as grounded in a reductive explanatory hierarchy organizing all the sciences; (2) a further belief that the manifest image or common sense framework somehow belongs in that hierarchy. Neither of these assumptions is common coin any longer, and we do not need to adopt them ourselves. I have strong doubts that Sellars falls into the second error (in fact I don't think he thought that the manifest image was a "theory" in any interesting sense)—but he was well aware that the manifest image is not just another scientific theory. The rest of Brandom's chapter is an argument for his expressive pragmatic naturalism as the right development of Sellars' worthier thoughts. That may, in fact, be the case; it is not something I am going to dispute here.

But I now want to offer a different perspective on Sellars' story concerning the two images that I think puts it in a better light, and one that has some claim on Brandom's own credence. I happily agree with Brandom that Kant's distinction between the noumenal and the phenomenal was a model for Sellars, who pretty much says this explicitly. But I don't think it is the only relevant model, for two reasons. First, Kant's phenomena/noumena distinction is absolute, never to be overcome, even in principle; so things as they are in themselves remain forever and in principle beyond our ken, and of course Sellars explicitly rejects that view. Second, the relation between the manifest and scientific images is essentially historical and developmental. The scientific image is supposed to grow out of the manifest image; it has the manifest image as its

necessary condition, not just logically or epistemologically, but historically and methodologically, and develops certain aspects of the manifest image in ways that ultimately turn around to challenge that image. And looked at in this light, *don't we have to say that in Sellars' view the scientific image is the Aufhebung of the manifest image?*

So I propose to take seriously some Hegelian aspects of the MI/SI relation. Both Hegel and Sellars reject the absoluteness of Kant's distinction; both think that knowledge of things as they are in themselves is not in principle beyond our reach. Both think that the reason they can reject the idea of the *Ding an sich* is that we need not start from dualistic assumptions, but from the belief that minds and their objects are parts or aspects of a single reality. And both understand the history of humanity as a development toward an ever-more-adequate set of categories in terms of which we can get at the very being of things. Notice that the interpretations of the MI/SI relation that Brandom examines are themselves static. It is consistent with those interpretations that the two realms of senses or the discovery of the underlying realizers of functionally characterized items are historically and developmentally linked, but it is not *essential* to understanding either the sense/reference or the function/realizer view. Any historical or developmental relation between the MI and the SI seems merely contingent.

Furthermore, the relationships that Brandom has in view take seriously the idea that "everything we talk about in the manifest image that exists at all" needs to be specifiable in the language of ideal science. To someone in the MI, who assumes therefore that there is reason to preserve many of the objects and concepts of that framework, the MI has to appear as a "large-scale map of reality to which science brings a needle-point of detail and an elaborate technique of map-reading" (PSIM ¶21; in SPR: 8; in ISR: 376)—remember that was one of Sellars' characterizations of perennial philosophy—a view Sellars explicitly rejects. Sellars' rejection of such a construal of the MI/SI relation might seem to support Brandom's criticism, because reductionism cannot construe the MI/SI as a filling-in of detail. But we need not interpret Sellars as holding that either we can give a determinate reconstruction in the language of ideal science of manifest objects or those objects do not exist at all or are somehow illusory. Science, in Sellars' view, will not be simply a filling-in of the details concerning the world we live in, nor will it be a simple redefinition of manifest objects in the language of ideal science. It will challenge and almost certainly revise some of the fundamental architecture of the conceptual framework we use to cope with the world. Sellars gives us a glimpse of this when he suggests that the basic entities of future science will not be particulars, or at least not objects as we now think of them, but such things as absolute processes. (Are those things particulars? I guess so. And by the way, I want to confess: Sellars has two articles devoted to particulars, and the logic

of particulars, and I don't understand either of them! So if someone can tell me what the hell is really going on there—please, be my guest. They're beyond me. . . .). The differences between the framework of ideal science and our MI will not be minor. We have, perhaps, only a taste of how different they may be in such manifestly odd conceptions as the wave/particle duality or a superposition of states.

Brandom's arguments turn on the difficulty of finding some relation between sortal concepts of the MI and sortal concepts of the SI that preserves the modally rich structure of manifest image concepts. The difficulty of doing so disrupts claims to a simple identity between the objects of the MI and the SI. Despite his talk of such identities in PSIM (and he *does* use that verbiage systematically there), it is surely Sellars' considered view that science will be developing successor concepts to those of the MI. The kind of developmental change Sellars has in mind cannot be a set of minor adjustments in a theory, leaving everything else, including the fundamental conception of the theory's objects, untouched. So just as consciousness and self-consciousness in Hegel's *Phenomenology of Spirit* discover new forms of objects as their experience unfolds—the experience changes, the object changes—there is every reason to believe that the conception of an object changes as science develops. In fact, if science proceeds as Sellars thinks it will, moving toward a pure process view of the world, the notion of at least the basic objects of our framework will change very significantly, for absolute processes apparently don't, for instance, belong to kinds. As I mention in my previous paper (in this volume), he wants us to get beyond kind talk somehow, get beyond causation. I doubt I understand how radically different the world looks from such a perspective.

The development of an adequately articulated SI will be a long and arduous process, spanning numerous scientific revolutions, revolts, paradigm shifts or intermediate frameworks. Sufficient structure from the old conceptual framework will have to be preserved at each juncture so that the new concept or concepts are reasonably seen as successors to some of the old concepts, but the successor relation here requires only relevant similarity, not identity. Over generations of conceptual change, therefore, it is thoroughly possible that the concepts to be found in a distant future science will as little resemble our current armory of sortal concepts as we resemble trilobites.

The argument Brandom models on Sellars' anti-phenomenalist argument clearly does point up the difficulty of so thoroughly replacing uses of MI concepts in the subjunctive conditionals that articulate the significant sortals we use in coping with the world that we no longer rely on the descriptive resources of the MI at all. And that surely cannot happen all at once, nor could it proceed by a series of clear identifications of the objects across the developmentally related frameworks. At least, I see no reason to think it should or could so proceed. So I tend

to think that Brandom does put his finger on an important point when he remarks that Sellars assumed a pretty strong unity of science thesis. But if we extract that commitment from the picture, I do not think that we thereby render the MI unassailable. If anything, it becomes easier to envision how a variety of sciences, each tuned to particular issues, might replace various aspects of the MI piecemeal, rather than necessarily waiting until there is grand unified science replacing it wholesale. Rather than a grand reduction of the objects of the MI to a new set of basic objects, we get localized and opportunistic explanations that we hope will be mutually consistent, but might not have any deeper unity. Over generations of scientific development, is it unreasonable to think that the concepts in the clauses of the relevant subjunctive conditionals that articulate the sortal concepts we use will themselves also be replaced with scientifically honed and reconstructed concepts?

I am aware, however, that in suggesting that science may end up more of a patchwork of locally profitable schemes than Sellars would ever have countenanced, I am also betraying my original Hegelian inspiration. For Hegel was a grand unifier if ever there was one. (I mean, that I take to be obvious.) Sorry, I can only say: so much the worse, then, for both Sellars and Hegel. Color me some shade of post-modern, I guess. I think I've given up on grand unities.

There is, however, another reason I claim Hegelian inspiration here. In my view, the Hegelian dialectic moves from the abstract toward ever-increasing concreteness, which I take here to be an ever-increasing comprehension of the context or whole within which the original abstraction is intelligible. It is no historical accident that the earliest developed sciences are highly abstract, and that progress in the sciences has often been achieved by beginning from a relatively simple abstract idealization and developing a theory of a larger context—a higher unity, one might say—within which the simpler abstraction becomes intelligible. It is in this context that I would see, for example, Newton's unification of Galileo's laws of terrestrial motion with Kepler's laws of celestial motion. Darwin provides a context within which a wide range of previously understood but "smaller-scale" phenomena—from breeding patterns in plants and animals to the newly won understanding of geological time scales to the organized diversity of species—fit together into a systematic view of biological phenomena. So it is just as much a Hegelian point that once the theory (or concept) of the larger context or higher unity is developed, the previous, partial, and abstract theories or concepts are not left as they were. Newton did not leave Kepler and Galileo untouched; Einstein did not leave Newton and Maxwell untouched; and Darwin left everyone in his field scrambling to rethink the phenomena they studied.

The objects cannot reveal themselves entirely in tightly constrained contexts in which they are abstracted from their normal situation. As

we broaden the context against which we see them, new and often hitherto unidentified aspects, properties, or relations come to light. Our very concept of objecthood changes. Did Sellars think that the progress of the sciences toward some Peircean ideal promised us a sequence of ever-more-adequate conceptualizations of objecthood itself, eventually swamping the hoary old manifest image concept of an object and the relevant categorized sortals?

The stumbling block I see to the thoroughgoing supersession of MI concepts by SI concepts is, as I have argued elsewhere, the fact that any conceptual scheme we could use must be *one that we can use*. We are finite beings with limited intelligence and built-in computational power, restricted to certain modes of sensory access to the world, however much we supplement them with instrumentation. We are necessarily located in space and time, we have certain natural needs and desires, and we are always operating within a cultural context that determines both our further interests and our further cognitive powers. Perhaps most basic is the fact that we must always be able to engage the world from the perspective of the singular human individual subject, who is also a singular human agent. In this sense, there is a privileged context, and it is not the "view from nowhere" context that, arguably, science strives for. There is no easy accommodation of this perspective in the sciences. And this is why Sellars thought the language of individual and community intentions must be joined to (or I would prefer, rather, never abandoned by) the vocabulary of science. Thus, I am inclined to think that something like the subject naturalism that Brandom, following Price, endorses at the *end* of his chapter is closer to the truth than object naturalism, but as I have tried to argue here, Brandom needs more argument really to get us there.

JIM O'SHEA: I am in agreement with most of what is contained in this powerful book. In particular, I find Brandom's extended defense of his "Kant-Sellars thesis about modality," which is elaborated throughout the central chapters 3 to 6, the constructive heart of the book, to be both innovative and yet interpretively compelling with respect to the real semantic heart of Sellars' philosophy. I myself agree that one of the deepest of insights elaborated in different ways by Kant, C. I. Lewis, Sellars, and now Brandom is that the objective purport of any empirical concept presupposes the prescription of lawful modal constraints governing the objects themselves, as systematically reflected in our inferential practices. This is Kant's idea of the categories as concepts that prescribe laws to appearances, and Sellars' idea of concepts as involving laws, and C.I. Lewis defends this, too—and now I think that Bob has really developed that in a nice way. And I also agree that these sort of normative presuppositional relationships, in general, constitute a key strand running throughout Sellars' philosophy: Sellars' idea, for example, that the normative espousal of principles is reflected in uniformities of practice

and in certain natural regularities. (I tried to make that the center of my book on Sellars as well, in not nearly as sophisticated a way, however.) This highlights the pervasive Janus-faced interplay between what is explicitly asserted on the one hand, and the various normative practices and natural regularities that are thereby, Sellars will say, presupposed or conveyed or prescribed on the other.

Brandom's book, however, theoretically analyzes and probes these matters more deeply than any other work on Sellars, providing a fully worked out theory of how the normative, the modal, the semantic/intentional, and the categorial-ontological domains of discourse stand in various complex relations rendered systematic by Brandom's conception of pragmatically explicitating meta-vocabularies. So it follows on from *Between Saying and Doing* in a very ambitious and interesting way. And he shows how all of this was anticipated with much (if partial) success in Sellars' own views, not to mention in certain ways, those coming out of Kant.

Furthermore (and I didn't do this in my book on Sellars, because I was mostly clarifying what Sellars thought), I also agree with Brandom's central critical recommendation that we should reject Sellars' only quasi-Kantian contention that the object-ontology of the manifest image is strictly speaking false. In particular, I am inclined to think (more along the lines of a Paul Coates or a David Rosenthal, or other theories of perception) that a Sellarsian critical realist theory of perception can embrace a richly explanatory theoretical posit of nonconceptual sensory representations while rejecting Sellars' implausible homogeneity argument and his disputable qualia intuitions (in the end he wouldn't call them 'qualia intuitions' in that way), thus rejecting Sellars' bravely held Feyerabendian idea that our manifest perceptual ontology rests not on various misconceptions but rather on one big, global locational mistake.

So overall, then, I agree with Brandom's general attitude toward each of the two Big Ideas from Kant around which he structures Sellars' philosophy: namely, the Good Idea of pragmatically elaborating Kant's pure categories as functional and explicitating meta-concepts; and the Bad Idea of warping—Sellars *knows* he's warping, he says this—Kant's phenomena/noumena distinction in the attempt to map it onto Sellars' own global distinction between the manifest and scientific images, and in particular to fit the 'global mistake' picture of Prichard: Perception rests on a mistake. So I agree that we should reject what I will call Sellars' *noumenal scientific naturalism*, as putatively entailing the strict falsity of the object-ontology of the manifest image. But now this provides a good transition to some divergences, at least in emphasis, in our readings. For I think we should reject Sellars' noumenal scientific naturalism but defend what I will call Sellars' *integrated* scientific naturalism, where the latter involves far more than just the sensible rejection

of instrumentalism about theoretical entities, and more importantly, results in a different portrait of the enduring significance of Sellars' philosophy. While I embrace quite a bit of what Bob also says is of enduring significance.

The following points have also been discussed very nicely by Dionysis (in this volume), and Bill (in this volume). Consider, first, Brandom's rejection in chapters 1 and 6 of what he calls "strongly cross-sortal identity claims," as "requiring that when manifest-image expressions refer at all, they must refer to items referred to by expressions belonging to the scientific image" (Brandom, 27). Without getting into the details of his very sophisticated argument, I can say that since I agree with Brandom's modal Kant-Sellars thesis, I am also inclined to agree with the specific anti-reductionist points he makes here. However, I don't think Sellars' scientific realism was intended to require strongly cross-sortal identities in that way. To bring this out I'll note that Bill deVries had raised an interestingly related objection in his Sellars book when he argued that Sellars cannot intelligibly hold both that manifest objects are *appearances* of scientific realities and that they are nonetheless also *identical to* those realities. For the appearance relation is asymmetrical while identity is not.

But I replied (in my book, 2007: 159–60, 222*n*22) that Sellars' general account of conceptual change and theory succession eludes Bill's objection, and I think it eludes Bob's objection, for similar reasons. Consider Sellars' views on the counterpart concepts of Newtonian and Einsteinian *mass*, just to illustrate the relevant account of conceptual change and of co-reference (co-reference can be complicated, and I'll say something about that in a moment), and of approximate truth with strict falsity. So what are those concepts all about? Sellars is going to distinguish $concept_1$ from successor $concept_2$, for example Newton's *mass* from Einstein's, but he's also going to distinguish $refers_1$ from $refers_2$, and $sense_1$ from $sense_2$; he's not going to have the same notion of co-reference as the standard sense-reference account has it. That is, it's going to be dot-quoted names, embedded within dot-quoted sentences (senses), and it is all going to concern functional similarity. The straightforward identification of Newtonian mass with Einsteinian mass, given their modally diverging lawful entailments, will rightly be ruled out by Brandom's Kant-Sellars thesis—the straightforward identity of the two. But Sellars doesn't straightforwardly identify them in that way, and nor does he do the classic Putnam-Kripke preservation of reference, either. Rather, he argues, first, that there is sufficient comparative, intra-theoretical functional role similarity between the two concepts to regard them both generically as *mass concepts* that are involved in relevantly similar explanatory tasks. (Jay Rosenberg was good on some of this.) The Einsteinian theory then provides a model, using its own Einsteinian conceptual recourses, of the Newtonian concept of mass, thereby demonstrating, internally to Einstein's theory, the strict

falsity of the relevant modeled Newtonian laws, while also exhibiting their approximate truth (that's why it works within limits). The strictly speaking falsity and reference failure provide the aspect of non-identity (that is, there is a sense in which there *is no* such thing as Newtonian mass, strictly speaking; it has been replaced); whereas the generic functional similarity, along with the Einsteinian reconception of mass and the counterpart modeling within the later theory, warrant the claim that relativity theory has thereby identified what mass really is and always was. This is Sellars' account of theoretical identification by counterpart reconceptualization, and that's what he's essentially saying we do, and it is supposed to apply as well to those aspects of the manifest ontology that fit this model of explanation. (David picked out some passages from *Science and Metaphysics*—it's very complicated—what do you do with all aspects of the manifest image, do you identify them in that way? And Sellars tries to say some things about this, but it's very complicated. Sometimes you just want to say there aren't those things, but sometimes you want to say they are identifiable with the later things.) This account of theoretical identification by counterpart reconceptualization is supposed to apply as well to those aspects of the manifest ontology that fit that model. Our belief that the banana is yellow, on Sellars' view, which is perfectly true and reveals reality if that worked, but it contingently turns out to be the case that there is a better theory that is re-conceptualized in such a way [that] it turns out it rests on a mistake, although it needn't have, and is strictly speaking false (which is the part I don't go along with). But Sellars' thought was that an analogous counterpart concept of the banana-shaped expanse of yellow will be part of a sophisticated neurophysiological-cum-environmental successor theory that will explain the approximate truth and hence the 'appearances' of the manifest ontology.

This is my shot at it: Both the reductive identity and the appearance claims, I take it, are supposed to be accounted for in that sort of way; and as far as I can see this account would then just sidestep Bob's and Bill's objections, because you can account for the appearance along with the aspect of identity, and you don't make direct cross-sortal identifications. So this would be one respect in which I reject Sellars' *noumenal* scientific naturalism, the *global* Feyerabendian ambition, without rejecting a central methodological and ontological component of Sellars' strong scientific naturalism. That said, I hold, as do Brandom and deVries, that many objects of the manifest image and of the 'special sciences' (I'm interested in the case of biology) are simply not plausibly subsumable under that particular *replacement* model of explanation—although, who knows, the future might hold in store some surprising reductions of this kind in some higher-level domains. Darwin gave us a different model for domesticating biology. Roy Wood Sellars was very good on the biological outlook and the layered aspects of things, and I'm attracted to that in these domains. And having also rejected

the color argument for scientific noumenalism, I am now left with no motivation, or no strength, for the heroic Feyerabendian attempt to envision a global idealization of the replacement model of theoretical identification.

However, although I shy off that super-task, I also see no obstacle to *integrating* any such cases of theoretical replacement identification—no logical objections to that story—*within* a scientific naturalism that, in other respects, is multilayered and non-reductive. I also see no reason not to seek to expand such theoretical identifications as a kind of empirically open-ended regulative ideal in Kant's sense, though again my own view is that such an ideal is probably not best thought of as achievable globally, for the reasons cited. But there are also no *a priori* philosophical grounds to *resist* such enthusiastic expansions, I think, in the object-domain either, wherever they might turn out to be plausible. I think, in a sense, the single most important and successful aim of Sellars' whole philosophy was to show how even the ideal Feyerabendian achievement of such an aim, if it really were coherently envisionable, would not threaten our autonomy as knowers and as rational agents. That was his biggest thing. That was his heart of hearts. He had two hearts of hearts, actually. The other one was the expanse of yellow. So doing those two things took some work.

Here is a further point: There are many other significant ways in which a slightly different overall picture of Sellars emerges once one disentangles noumenal scientific naturalism from various more plausible but nonetheless still strongly naturalistic and ambitious aspects of Sellars' views—views that can, however, I think be fully integrated without replacing the manifest ontology. So consider Brandom's remarks in his introduction on his discussions with Sellars (fascinating biographical discussions, very revealing philosophically):

> how . . . matter-of-factual picturing relations [are] to be related to the normatively characterizable discursive practices that (he and I agreed) alone deserved to be thought of in genuinely *semantic* terms. [. . .] [O'Shea aside: How we going to characterize the matter-of-factual picturing relations, how are they to be related to the normatively characterizable discursive practices that] he and I agreed, alone deserve to be thought of in genuinely sematic terms. We both saw that it is in a story about how sign-designs can lead a double life, on the one hand as items caught up in a web of causal relations supporting subjunctively robust conditionals, and on the other as normatively characterizable as having proper and improper uses ("according to rules" . . .) that [Sellars'] response must be found. . . .
>
> (Brandom, 13)

(Brandom on his discussions with Sellars.) But Brandom reports having been "quite critical of [Sellars'] characterization of this

amphibiousness [O'Shea aside: this double life] without having positive suggestions as to how one might better conceive it" (13). He says, however, that "the issue has come to assume an importance for me of the same magnitude as it did for Sellars," and he refers in this context to his key conceptions of "the normative and the modal Kant-Sellars theses," and thus to "deontic normative vocabulary and alethic modal vocabulary as articulating two aspects of the phenomenon of intentionality" (13). So we've got Sellars bangin' on about picturing in his office, and they're both agreeing "How are we going to relate this causal dimension to the normatively characterizable semantic dimension?" And then Bob later develops what I think is great stuff in the Kant-Sellars thesis about modality, and about normativity, and these show how causal modality is layered with normative vocabulary and with intentionality as well.

But I think it's clear that there are crucial differences between those two envisioned ways—Brandom's and Sellars' in these conversations—in which simultaneously normative and causal 'double lives' can be led by various kinds of tokenings, as they had put it: *both* sorts of 'double life' are in Sellars—both Brandom's and the other one—but only one of them is in Bob's work. Brandom's double life is a sophisticated pragmatic metalinguistic development of Sellars' conception of how "the language of modality is . . . a 'transposed' language of norms" (Sellars IM V, §39), which for Brandom is the semantic resultant of two pragmatically mediated metavocabularies. And that's good stuff. But in their unhappy discussions specifically about 'picturing', what Sellars was after—and I think Brandom knows this—was a different sort of Janus-faced relation between the normative and the causal, one that has nothing *essentially* to do with *noumenal* scientific realism at all, but has everything to do with the possibility of an *integrated* and *non*-reductive scientific naturalism. What Sellars had in mind is that for any physical tokening (whether it be a non-language using animal brain state, or a conceptual-linguistic tokening proper) for any such thing to have any object-representational cognitive content or purport in relation to its environment at all, such a tokening must be embedded within some wider normative system of proper functioning—whether it be a social-linguistic space or a naturally selected space of proper biological functioning. You're only going to represent something if you're embedded in one of these kinds of wider, normative proper functioning systems. The result being that it now intelligibly *ought to be* the case that representing-events—little tokenings of these kinds—having certain properties and relations are tokened *ceteris paribus* when and only when they stand in appropriate causal-historical-sociological relations to what are thereby *represented* environmental events that are supposed to have certain corresponding properties and relations (cf. O'Shea 2007: 147–158). This is essentially [Huw Price's notion of]

e-representation embedded within i-representation; Sellars is looking for something like that.

Roughly put, Sellars' thought was that there must be a way of embedding an underlying and norm-parasitic naturalistic theory of object-representational mapping and tracking *within*, and resulting from, in part, the familiar non-reductive, rule-governed norms pertaining to 'meaning as use' that both generate and depend upon those complex causal-historical relationships. It's an interdependence. There are familiar objections to such subsequently highly influential naturalistic strategies of accounting for representational purport in certain matter-of-factual domains—but usually precisely because they fail properly to appreciate what was Sellars' main lesson: the normatively Janus-faced character of the relevant naturalistic relations. You're not going to have purely naturalistic theories of representation except as embedded in one of those two kinds of wider normative spaces: proper biological functioning, or the space of reasons. Sellars' characteristic philosophical genius, for his part, was to contend only for the *intelligible logical space* for such substantive and integrated naturalistic hypotheses, as representing no threat to (and in fact requiring, in our case) the sorts of normative-inferential conceptual relations in terms of which our intentionality and our rational agency are constituted—and that latter stuff includes, of course, the sorts of meta-level pragmatic 'double lives' that Brandom is doing so much to clarify and expand. Sellars wanted that kind of transpositional meta-vocabulary 'double life', and also this other 'double life' where we're tracking the world as a result, but in a way that is parasitic on those normative activities and doesn't swing free in a sort of Kripke-Putnam referential way.

Again this sort of integrated scientific naturalism has nothing to do *specifically* with Sellars' overreaching noumenal scientific naturalism or with the conjectured wholesale theoretical replacement of the manifest image object-ontology with that of Peircean science. For Sellars' view—and this is a deep and important thing, that I think is definitely right—is that these Janus-faced norm/nature double lives pertain to *any* matter-of-factual discourse, including that of the manifest image. But unfortunately this groundbreaking and intelligible dimension of a normatively integrated scientific naturalism, which appears throughout Sellars' philosophy, seems unnecessarily—at least gives the appearance—of unnecessarily going out with the noumenal bathwater in Brandom's book on Sellars. The result is that Brandom's Sellarsian pragmatism gives at least the appearance of being more hostile to scientific naturalism than it needs to be, in my view, and thereby misses out on the chance to recover, in a non-reductive but enthusiastic way, one of the most dominant themes in the history of American pragmatism.

In this final section I do want to say a little bit about nominalism, but not much. Finally, then, a few words about Brandom's carefully constructed

critique of Sellars' nominalism in the final chapter 7, entitled 'Sellars' Metalinguistic Expressivist Nominalism'. Brandom argues plausibly—I think plausibly—that Sellars ultimately confused what should properly be distinguished as *pragmatic* and *semantic* metalanguages, as he was just explaining that to us helpfully in his opening remarks above (and Chapter 12 of this volume), and which were already distinguished to some extent in Sellars' own conception of a 'pure pragmatics'. The result of not distinguishing these pragmatic and semantic metalanguages carefully, Brandom argues, was that "Sellars himself draws invidious nominalistic ontological conclusions" (28) according to which only *nameables* (ultimately, 'pure processes') exist "in the world", as Sellars puts it (cf. 269). This view also brought in its train Sellars' lingering non-realist, empiricist attitude toward the alethic modalities, which is a nice connection he makes, and which is true in Sellars—I'm going to concentrate on that case. Brandom's recommended emphasis on Sellars' *pragmatic expressivist* metalinguistic account of the function of categorial vocabulary—'property', 'fact', and so on—pragmatic expressivist, not semantic metalinguistic, seeks to preserve key aspects of Sellars' Carnapian metalinguistic account of universals, properties, and kinds, but in a way that entails no revisionary ontological nominalism and no resulting hostility to modal realism understood in Brandom's sense.

I think this is a very powerful argument. This chapter insightfully diagnoses what is indeed, it seems to me, a genuinely deep and difficult choice-point in both the pragmatist tradition and in the Sellarsian legacy. First, I agree that Kant himself was a 'modal realist'—I quite like Kant, I work on Kant—in *something* very much like Brandom's sense: for on Kant's view the upshot is that we necessarily represent that necessary causal connections obtain in nature itself, objectively. (Nature itself—this has nothing to do with Kant's noumena; it's *nature*.) But Sellars' view was subtly different, and also compelling, and I think he held the following views with respect to *all* matter-of-factual domains (we don't need to bring in the noumenal anything!), including the manifest image, since he held that picturing-representations are generated in any empirical, matter-of-factual domain. So again: leave aside the noumenalism. Rather, it's the Janus-faced picture again: in this case (i.e., of nominalism with respect to necessary connections—the modal realism issue), Sellars' view has to do with our espousal of the normative Kantian inferences that thereby conceptually represent objectively necessary connections (the Kantian view is that to experience any event you have to judge that something preceded that event and necessitated it). And what's really nice about Bob's Kant-Sellars thesis about modality—and I think it is also in Kant and Lewis, and in Sellars—is that to cognize any particular event or object, you're already conceptually embedding within those objectively real causal connections. Or if it's not causal, it's modal in some other way.)

The key further point is that our espousal of the normative Kantian inferences that thereby represent objectively necessary connections, is reflected for Sellars in underlying naturalistic patterns of representing-events that track and map those Humean successions of events that *are all that is really taking place, ontologically, in the natural world.* That is, we're inference-ticketing ourselves and others, we're saying "if you assert this, then [you have permission to] assert that," and what we're doing is getting our inferences to go in certain patterns, according to normative rules that we espouse, we are representing that there are necessary connections (as it were), we are saying 'A necessitates B', but what those inferential practices are doing is mapping—in Price's e-representation way—the Humean successions that are *really* all that's out there, for Sellars. So we're getting ourselves, like Sellars' robot in "Being and Being Known" to say, to represent "lightning now; thunder soon", and we're doing that with fancy inferential norms, and with ascriptions of causal connection, but what we're doing is tracking nature. Sellars wants to say *in nature*, there is a sense in which Hume was right. There aren't As followed by Bs *and necessitations*, and representing the latter in a way that tracks them. And that's an interesting idea. This is why I find it an interesting and difficult choice-point that Bob is really picking up on.

So Sellars' doctrine of 'pure processes' and fancy later stuff are just more adequate replacement versions of this basic story, which he already tells in "Truth and Correspondence" *about the manifest world.* He says there is an underlying Humean truth, which our normative Kantian practices, as it were, are getting us to track. That's the story, and it's a lot like some things Huw Price was saying (in this volume). So Sellars' constitutive Kantianism, in other words, is supposed to be consistent with this underlying Humean or Russellian view about what's *really* contained in nature, which doesn't include 'necessitations' in the same way that it includes objects—I mean, Bob Brandom can say it's not the same way, too, and can make distinctions here, but for Sellars in a stronger sense, they're not in the world in the way that objects, events, or 'pure processes' are. So the good thing about this Peircean science, we're doing all this normative inferring, and then we're reacting to apparent disconfirmations, we're changing our inference-tickets, the sorts of laws we espouse in response to those things, and all of this explanatory or explanationist project is getting us to make causal inferences, reflected in inference-ticket licenses, and they are getting us to better track the world. That's what science is doing—it's changing our inferential norms to ultimately be better tracking what's really out there, which is one damn thing after another. And then one damn thing after another can be conceived in more fancy and more fancy science. So that's Sellars' picture. And *that's* an interesting picture.

So I'll end with this: do we go with the modal realism of Kant and Hegel and Brandom on *this* question, about necessary connections, or do we go with the underlying naturalistic nominalism of Sellars' corrected Kantian empiricism? One thing I know is that Sellars' *noumenal* scientific naturalism, which is the focus of Bob's criticism, is irrelevant to that debate, although it is maybe connected in the diagnostic way that Brandom just talked about in his opening remarks. I suspect that an integrated scientific naturalism can, perhaps, contrary to Brandom, retain the merits of Kant's objectivist view, that we represent that there are objectively necessary connections, while also preserving Sellars' view that, considered from a certain ontological point of view, universals, real connections, and other abstract entities don't exist in nature, *per se*, at the end of the day. For now what I can say is that I think Brandom's critique of Sellars' nominalism, whether successful or not, has succeeding in highlighting the importance of this particular fork in the road, very nicely.[1]

I probably have another minute, do I?

MICHAEL HICKS: One more minute!

JIM O'SHEA: One more minute! Perfect. When one says there are no properties in the world it can sound very implausible, but one of the things Sellars does in those particular articles is say that what's really out there are *qualified particulars*. There are, as it were, propertied objects, a kind of trope theory is what Sellars' ultimate ontology is. When you are using the word 'red' as a dispensable auxiliary predicate, and you're saying "X is red", the way we use that does then succeed in getting you to track red objects and not green objects. In a sense what Sellars thought Plato's abstract entities were doing in carving things at the joints, they were really carving our representational activities at the joints by showing what kind of predicate practices we have that enable us to track objects in this way. Sellars' theory was that we could do it without predicates, and we can also explain the upshot of it just in terms of, as it were, trope-like objects—qualified objects without any further universals, and so on. So he's got some interesting things to say that don't make it sound so bare. And his later "Towards a Theory of Predication" offers a possible account of what that might look like. I'll end there. Thanks.

BOB BRANDOM: Thank you both for that. I really don't disagree with what you've said. Jim is characteristically judicious. I think we all agree in rejecting what Jim is calling now the *noumenal scientific naturalism,* and I think we all agree that this really is a strand that is in Sellars that is mixed up with these strands that he didn't fully see how to disentangle, and we're now in a position to do that. I didn't take my argument to do more than—I don't know—drive a stake through the heart of *that* aspect of his rhetoric. That clears the ground for us to

think about what is recoverable in that way. Some revisionary things are going to need to be done with what's left.

Bill, you started off giving us this quote about Sellars' seeing at the heart of his suggestion about the manifest image and the scientific image that the unification of them couldn't happen piecemeal—it had to happen sort of "all at once." And that's something that your attractive Hegelian picture says, "Well, that can't be right. It is going to have to be spread out, and piecemeal, and so on," and that's a very sensible way of going. At this point we have to decide which pieces we can use to make something sensible out of. And Jim, I think your account had the same general shape. You talked about this methodological fork in the road. I'm not myself tempted by the naturalist/empiricist strand that I agree is in American Pragmatism—but one might be—and seeing that it's going to be real hard to have both of these, we need to think about what considerations speak for the one and for the other, and which other desirable things do they integrate with? I think that shows us in a good place in trying to think critically about Wilfrid's work. And I hope it's obvious to everybody not only how sophisticated and intricate Sellars' view is, but just how rich the questions that remain are after we've engaged critically with some parts of it. It really seems for me a good way to end in our thinking about Sellars in a new generation.

BILL DEVRIES: I agree very much with Bob that one of the nice things about people who are interested in Sellars is that, unlike devotees of *some* philosophers, they don't think that philosophy *ended* with him. If anything, they think, "Oh, it's just getting going!", and I like that. Sellars is an invitation to do more, not to stop where you are and spend your life explicating something that's already finished. I'm with Bob 100 percent on that.

MICHAEL HICKS: We can open it to the floor now, if anybody else has questions.

KEVIN FINK: Thanks. This question is primarily for Jim. I like the very nicely nuanced discussion and defense of Sellars' scientific naturalism. My question is this: It seems like one of the main reasons to want to defend a Humean ontology, of the sort you were seeming to associate, would be that you bought into the Humean epistemology.

JIM O'SHEA: No, that's not true at all.

KEVIN FINK: Right! And this is sort of an oversimplification of a lesson that, at least I read out, from Professor Brandom's book, that once I *reject* that Humean epistemology, once I cease affording this privileged status to empirical descriptive vocabulary, when I realize that in order to have any of that in any meaningful way I must already possess these other (at least) abilities, if not concepts. It seems like you were conceding, on the epistemological end.

JIM O'SHEA: No that was the great thing [about Sellars] is to make none of those Humean epistemological assumptions—that's my view, anyway.

KEVIN FINK: You were conceding the Kant-Sellars thesis, it seemed.

JIM O'SHEA: Oh, I concede that.

KEVIN FINK: Yes, okay. You agreed that in order to have this empirical descriptive vocabulary we needed to have these other abilities, involving modality, etc.

JIM O'SHEA: Yes.

KEVIN FINK: So I guess the question is, given *that*, given that the empirical descriptive vocabulary is not privileged in that sense, why privilege it in the ontological sense? Why think the ontological picture is just about the things that we're supposedly describing, and not about the modal facts, etc.?

JIM O'SHEA: Well, I guess I wouldn't privilege it . . . It's in a different line of work. But it is unique; think of how bold Sellars' attempt is to combine the later Wittgenstein with the early Wittgenstein. The way he's doing it, he's reading the early Wittgenstein purely as a theory of mapping, picturing, and then saying we can have *both*. Does he then *privilege* the later Tractarian account over what was absolutely central to him to the end—the whole later Wittgensteinian picture? I don't want to put it *that* way. With regard to certain tasks it does a different thing. In the way that Huw talks about these—it's in a different line of work. We can explicate that line of work, but it is still embedded—it doesn't have to be *privileged* in the way of being world-revealing, because moral claims can be true and false, too. And—what's the quote that keeps coming up?—an "ungrudging recognition" of all the other ways in which we say true things about the world doesn't necessarily have to be made second-class, if it turns out that there is a useful distinction of this kind. Maybe Sellars fell into making that [distinction] too strongly when he's a logical atomist at heart—when he says "I'm a Russellian"—all those things that you'd never believe that he'd say if you'd read his epistemology.

BILL DEVRIES: I still think that *he* thought—I don't know rightly or wrongly—that the move to picturing did give some kind of privilege to those entities. Because, in his view, if there were no picturing, nothing else would work, because nothing else would get hooked up properly to the world in which we live. It's the picturing that ties us to the world in which we live, so that gets priority because it is the *sine qua non* for everything else making sense. So all the rest of the stuff is not *second class*, and we can ungrudgingly recognize that it's *different*, but I think that he does think that there's some privilege to the vocabulary in which one pictures. But he doesn't think it evident what that vocabulary is. He thinks that every language has to have picturing, and pictures don't make themselves self-evident to us *as* pictures. That's a scientific project: to figure out what it is, and how it is, that our language, either here-now or in some future scientific millennium, pictures, and what it is about that language that pictures.

CARL SACHS: I just want to raise the following consideration to see what the feeling in the room is about this. One of the problems you're going to have with Sellars is his commitment to the Humeanism. I think that is actually going to be a very serious problem, in a lot of ways, and the person you need to make the Sellarsian project work is not Hume, but Dewey, because it's in the organism-environment transactions where the rubber of discursive practices really hits the road. I'd like to invite some comment on that, or maybe make some push-back: What's wrong with using Dewey to do this? Thank you.

BILL DEVRIES: Let me ask for further clarification. What's wrong with Hume? It's the minimal, right? The world is just one damn thing after another. That's the minimal, I assume. I cannot think of anything more minimal than that. So why do we have to think there's more?

CARL SACHS: That's funny!

BILL DEVRIES: *That's* pushback!

CARL SACHS: That's fine—I asked for it! So the question is: Why more than just what Hume gives us? Well for one thing, we might want a picture that is post-Darwinian, informed by better biology, more consistent with what we do know in 2015 about how brains work, and how ecosystems work, and things like that, so that's one point in favor of it. Another is that because it is more richly structured, because there is a lot more going on, it gives us more information about *how* the higher-order discursive practices get constrained. We don't have the problem of our concepts just being projected onto the world—a cookie-cutter picture of how concepts cut into the world—which might be a worry with a Humean background.

BILL DEVRIES: (pointing across the room) Huw!

MICHAEL HICKS: Huw, do you want to answer that?

HUW PRICE: Well, it's sort of related to this discussion. As I've been listening to this I've been asking myself why is it that in some respects I'm a Humean or a Russellian about causation. Causation is a topic on which, in one sense, I have developed views that are in some sense independent of these broader issues that we're talking about. I think [this] is a case in which a good case can be made that we can understand our ways of causal thinking by combining a Russellian or Humean ontological picture with the fact that we're *agents*. By putting the two together you get the bit that seems to be missing from the Russellian picture—the bit that Nancy Cartwright rightly pointed out needed to be there in order to account for the distinction between effective and ineffective strategies, but I think wrongly thought had to be there in the metaphysics. She didn't see the possibility that you could understand it in terms of the distinctive standpoint of an agent. But then that thought led me to two thoughts. One is: To what extent whether what you need to put into the explanatory base can be used more widely to gain traction on the sort of more general issues we've been considering here. But then also to a

suggestion about how we can modify what Bob has given us here, with these diagrams, to make those questions explicit. And that's by modifying them, to put in the P-boxes, to make explicit what you're assuming about the world in which those practices are taking place. In a sense, I suppose, it must be implicit, anyway, but it would be nice to make it explicit. Then, for example, in the causal case, you could have in the P-box a sort of Humean or Russellian world of one thing after another, just regularities and associations, and then you put in an agent—that's the sort of P-bit. My claim would then be that at that point you have what I'll call POV—practice and ontology—sufficient for the causal vocabulary. So this is really not a question; it's a suggestion as to how we could use this material that Bob has already given us, and the very nice graphical representations, and add to it this ontological element. The question is, *what's assumed on the world side* in these explanations and vocabularies? That might make some of these questions we're looking at a little bit more tractable.

BOB BRANDOM: Well I'm down with that. One way of thinking about the advantage of thinking about the relations of discursive practices to things in the world is that you're not tempted to think, "Here's the word, just sort of sitting still, and here's the practice. Now how are we going to get these together?" If your paradigm of a practice is attaching two things by driving nails with a hammer through them, the hammer and the nails are *part of the practice*. They aren't sitting over here while the practice, or just the movements of your arm, without the things in them. The way it is supposed to work with those diagrams is that what's in the practice is always a matter of the vocabulary you're using to specify the practice. Of course to specify hammering, or discursive practice, one is going to be characterizing the objects that are involved in the practice. So I would say we're just going to get one more box—that's the vocabulary we use to specify it—but yes, there is a big difference between using a richer vocabulary that's describing the abilities, the roles the objects are playing, and so on. And I think it would be an interesting task to translate your account of causation into such a specification of the practices.

BILL DEVRIES: I want to go back to Carl, because I don't want him to think we just ignored him.

HUW PRICE: Oh, I'm sorry, I had meant to say how I thought it connected to Carl's question. Apologies, Carl!

BILL DEVRIES: Sellars never really defends this [Humean picture]. He just assumes that when it comes right down to it the world is just one damn thing after another. There's no defense of it. If I were to construct one in his name—which is always a dangerous thing to do—I guess it would be something like this. He does recognize, of course, that our conceptual frameworks are multidimensional, highly complex, there is a lot of richness with modalities, and normativity, and all those other things, but as

Stephanie [Dach] pointed out, I don't believe there is an inference from "Our conceptual framework has this rich structure" to "the world must have this rich structure," because conceptual frameworks are changeable, discardable things. So there's no straightforward inference. The thing that I think he holds on to as sort of the last ditch is just the picturing. That's the ultimate standard. And when we finally get a conceptual framework that allows us to picture arbitrarily well arbitrarily designated segments of the world history (or whatever)—at *that* point we have some right perhaps to assume that the structures of our conceptual frameworks adequately, in some way, represent the structure of the world. But actually, I'm not convinced. At that point, the picturing is what we've got. I'm going to make *minimal* assumptions about what the hell's out there, except insofar as we can build a framework that enables us, in the long run, to construct these pictures that really do work. Everything else is on *our* side.

BOB BRANDOM: You're bringing me back to those discussions with Sellars. All we need to make this work is an Archimedean point outside all conceptualization, a point of view from which the goodness of all of the different conceptualizations can be compared, so we can be confident that we can really be in touch with the world.

BILL DEVRIES: The trouble is, there's no such point of view!

BOB BRANDOM: Exactly! Right! The way I think about picturing is, it's an inferential matter. To say that the one thing pictures another like a map is to say that inferences from a range of map-facts to a range of terrain-facts are good. So any assessment of a picturing relation is done from within a language in which you can specify the pictured facts and the picturing facts so that you know, well, from this design on a map—that's a map-fact—and you can make an inference to this one to another one—no, that's a coffee stain, there is no good inference from that to the way things are. *That* I can understand. Those relations are important. But they're precisely *not* that Archimedean point, that language that Sellars uses. . . . I don't see that that's intelligible.

BILL DEVRIES: But at least you said I sounded like him. I was just trying to sound like him!

JIM O'SHEA: If Sellars sounds like that in some context, he wants neither an Archimedean point that's outside of all frameworks, or to have using maps be using them in that agential way. What he thinks is that through the use of discursive practices and everything governed by norms, in the normal way that would be accessible to us, let's say that we're telling each other to infer Q, after asserting P. Let's take the robot where it's two dots when it's lightning, and it's '//' when it's thunder; we engineer this into the system, and it's our engineering knowledge that is the normative practice that sets up the [picturing relations]. . .—just like record companies engineer all the intelligible inferences that you can make. But we can also see there is an extensional relationship, is his idea, between

these dots, and they occur in a causal historical relationships, (but) of course we engineered that. Now, our discursive practices are the only things we have access to, and falsification—we change our theory—all of it has got be internal to a framework for Sellars—he believes that—but I think he thinks our actual tokenings as a result of those (and this is the part, I don't know if there is a good theory for this) are unwittingly, as it were, setting themselves up in various temporal and spatial structures that are mapping things. So I don't think he has the "user-map" view of pictures, he's got the underlying—

BILL DEVRIES: They are like animal representations.

BOB BRANDOM: That's absolutely fine. What I'm asking is, what are you *saying* when you say there is this extensional relationship between the things? It's that *there's a good inference* from inscription facts to meteorological facts.

JIM O'SHEA: He says it is the sort of thing that can be articulated by a mathematical theory of second-order isomorphisms—a similarity between the relationships among the dots on the record and there's a second-order isomorphism where they can be shown to track mathematically the . . . But it all results not from magic, but from the rule-governed practices.

BOB BRANDOM: I wasn't addressing sort of how it's established—just what you're saying when you're saying there's this picturing.

PRESTON STOVALL: That's what you're *doing* when you're saying there's picturing—you're saying here's an inference ticket schema from this set things [of] in the nervous system to this set of things in the world. But what you're *saying* is that there's a subjunctive correlation between this nervous system and this set of events that this kind of creature has the capacity to be on to. So what you're *doing* is you're underwriting inferences. But what you're *saying* is that there is a correlation between two structures in the world that's isomorphic in virtue of a subjunctive correlation between them.

Sorry, I had a question, anyway. I wanted to say something in defense of Carl, and precisely along this point. [There are] two ways in which Dewey might be useful here, both of them connected to Sellars. These are things that I think Sellars picks up from Dewey. One is emphasis on evolved habituation. So it's in Peirce too, and obviously in James as well. But particularly Dewey's focus on reflex actions, and states of the central nervous system as conditions on the possibility for doing the kind of things that we do. You might think that, insofar as this story that you suggested Bob, on how the social and the natural might be united—getting the son and the daughter together again—I think Dewey is someone, because he thought seriously about the kind of things Sellars was thinking about, might be useful in trying to give that reconciliation. The second thing is just a passage that's thick in the middle of *Experience and Nature*, and it reads like an abstract of "Empiricism

and Philosophy of Mind" because it asserts both a rejection of mythical givenness, and circumscription to language. So he's got psychological nominalism here, in virtue of the phrase "apart from discourse." I'll just read—this is Dewey: "The notion that sensory affections discriminate themselves, apart from discourse, as being colors and sounds, etc., and thus constitute certain elementary modes of knowledge, even though it be only knowledge and their own existence, is inherently so absurd that it would never have occurred to anyone to entertain it, were it not for certain preconceptions about mind and knowledge."[2] So he's rejecting the thought that apart from discourse there could be these momentary sensory episodes that could constitute baseline units for knowledge. So again, I think there might be reasons to think, insofar as we're interested in Sellars in a new generation, we take seriously the roots in classical pragmatism and Dewey.

BOB BRANDOM: Sellars himself never called himself a pragmatist, but was very apologetic about that. There is this poignant passage where he says, "My father didn't really like the pragmatists"—they had displaced his generation of philosophers institutionally. Roy Wood Sellars is one of this generation of philosophers who used three names—Ralph Barton Perry, William Pepperell Montague, Edwin Bissell Holt—that had the critical realism, direct realism, representational realism debates in perception, all of which, when we look back at them now, we say "This is some of the same sort of stuff we're really worried about—this was a sophisticated discussion. Where did this all go?" It was swept away by enthusiasm for Dewey. So his father was actually quite resentful about pragmatism. Sellars said, sort of apologetically, "I can't really take them seriously in print, because of the way I was raised." [Laughter from the audience.] This is just a historical footnote. Yes, we can say we can see all these things, and there is a sense in which he suspected that was true, or knew that that was true, but just for contingent reasons, couldn't pursue that.

AUDIENCE MEMBER: I thought it was the other way around. I though he was distancing himself from pragmatism because it had been eclipsed by another philosophy. So thank you very much for that.

JIM O'SHEA: Just another Carl thing . . . If that sort of enactivist Deweyian thing were true, it might fit better with Bob's outlook of vehicle-less thought and not positing inner tracking items—it's much more a view of our environmental nature not requiring a special mode—it might fit well with some of Bob's project. And I'm happy if that turns out to be true.

MARK LEBAR: Yes, my question concerns the categories. I don't have a problem with the description you give of the purpose of it, and the role and the function of it, but I'm putting a different twist on the question, given the year we're in today. What role do the categories play? Do the

categories need to be revised in some way, as far as Kant goes? Could you address how that could affect the formula you give for how these functional roles are going to perform?

BOB BRANDOM: The list of concepts that have this status, although Sellars didn't call it categorial, playing this covertly metalinguistic expressive role that *I'm* saying was his notion of the categories was somewhat different from Kant's already, and *my* list would be different yet. So, for instance, Sellars was downstream from Wittgenstein's schooling Russell on the nonrepresentational character of logical vocabulary, where Russell really worried about negative facts, and conditional facts, and so on. Wittgenstein had taught everybody the lesson you don't have think about logical vocabularies playing *that* role. Sellars did not put logical vocabulary in a box with modal vocabulary, with ontological categorial vocabulary; he had a special box for that, and actually we hear very little about how he thought they worked. They're just special cases. But I think they belong in that box. Sellars, I think correctly, treats normative vocabulary as going in that box. Kant doesn't. As much as I think the First Critique is organized around that insight into the normative character of discursivity, that's not one of the categories. Now does it end up having that status by the end of the Second, and especially the Third Critique? Maybe. But for Sellars that's absolutely front-and-center, and I think seeing the intimate relation between the normative and the modal—the alethic and the deontic—that goes deep for him. Once you start thinking about these successor specifications of expressive roles that I'm saying are categorial, yeah, that's going to have an effect on the extension of what you say should be treated like this.

MICHAEL HICKS: Thank you very much. Should we give to the panel our applause?

(audience applause)

Notes

1 [Just for 'philosophical sport', as Quine says: an 'Afterword' from Peirce (not that it helps):

> "Kant was a nominalist; although his philosophy would have been rendered compacter, more consistent, and stronger if its author had taken up realism, as he certainly would have done if he had read Scotus. Hegel was a nominalist of realistic yearnings. I might continue the list much further. Thus, in one word, all modern philosophy of every sect has been nominalistic."
>
> (Peirce CP 1.19, from the "Lowell Lectures of 1903," Lecture IIIa)]

2 Dewey, *Experience and Nature* (LaSalle: Open Court Press, 1929), p. 212.

Contributors

Deborah R. Barnbaum, Kent State University. Deborah R. Barnbaum is professor and chair of Philosophy at Kent State University. She joined the faculty at Kent State University after receiving her PhD from the University of Massachusetts in 1996. She is co-author of *Research Ethics: Text and Readings* (with Michael Byron; Prentice hall, 2001) and author of *The Ethics of Autism: Among Them But Not of Them* (Indiana University Press, 2009), as well as numerous book chapters and journal articles in the field of bioethics. Not a Sellars scholar, she has nonetheless been dogged by Sellarsians most of her academic career.

Boris Brandhoff, University of Bonn, Germany. Boris Brandhoff studied philosophy, modern history and constitutional law at the University of Bonn, Germany, and the University of Lucerne, Switzerland. His dissertation was entitled *The Pragmatics of Representing in Wilfrid Sellars's Philosophy*. He received a doctoral scholarship of the German National Academic Foundation. From 2011 to 2014 he was an adjunct lecturer in the Philosophy Department, University of Bonn and was a visiting scholar at the Philosophy Department at the University of Pittsburgh in 2013. He is currently research assistant to the dean of the Faculty of Arts at the University of Bonn. He has given talks at international conferences in Dublin, Johannesburg, Prague, Kent Ohio, Beirut, and Erlangen.

Robert B. Brandom, University of Pittsburgh. Robert B. Brandom is distinguished professor of philosophy and fellow of the Center for Philosophy of Science at the University of Pittsburgh. He joined the faculty at Pitt immediately upon receipt of his PhD from Princeton in 1977, where his doctoral supervisors were Richard Rorty and David Lewis. He is the author of *Making It Explicit* (Harvard, 1994) and *Between Saying and Doing* (Oxford, 2008), and most recently *Wiedererrinerter Idealismus* (Suhrkamp, 2015), and *From Empiricism to Expressivism: Brandom Reads Sellars* (Harvard, 2015). He has also written eight other books and over one hundred articles. He is a fellow of the American Academy of Arts and Sciences, and recipient of the Distinguished Achievement in the Humanities Award from the A.W. Mellon Foundation and the Annaliese

Maier Forschungspreis from the Alexander von Humboldt Stiftung. He has delivered the John Locke lectures at Oxford, the Hempel lectures at Princeton, the Woodbridge lectures at Columbia, and the Townsend and Howison lectures at Berkeley. Website at www.pitt.edu/~rbrandom.

Dionysis Christias, University of Patras. Dionysis Christias received his PhD from the Department of History and Philosophy of Science at the University of Athens, and he is currently part-time faculty at the University of Patras. His research interests include epistemology, metaphilosophy, the debate between naturalism and normativism, and the philosophy of Wilifrid Sellars and post-Sellarsian philosophy. His recent publications include 'Can 'Ready-to-Hand' Normativity be Reconciled with the Scientific Image?' (forthcoming in *Philosophia*, 2016), 'Can Sellars' Argument be Used Against His Own *Scientia Mensura* Principle?' (forthcoming in *Synthese*, 2016), 'Sellars, the Myth of the Categorial Given and Correlationism' (forthcoming in *Journal of Philosophical Research*, 2016), 'Sellars *Contra* McDowell on Intuitional Content and the Myth of the Given' (*Philosophia*, 2015).

Willem A. deVries, University of New Hampshire. Willem A. deVries, University of New Hampshire, received his degrees from Haverford College and the University of Pittsburgh. He has taught at Amherst College, Harvard University, Tufts University, the University of Vienna, and UNH. His books include *Hegel's Theory of Mental Activity; Knowledge, Mind, and the Given: Reading Sellars's "Empiricism and the Philosophy of Mind";* and *Wilfrid Sellars*. Despite the focus on G. W. F. Hegel or Wilfrid Sellars, there are occasional forays into other territories, in a continuing attempt to figure out how it all hangs together. He is co-editor of Routledge Studies in American Philosophy and founding president of the Wilfrid Sellars Society.

Kevin Fink, University of South Florida. Kevin Fink is a PhD student in philosophy at the University of South Florida and visiting scholar in the Philosophy Department at the University of Pittsburgh. Kevin earned his BA and MA in philosophy from Kent State University where he early on became enamored with the philosophy of Wilfrid Sellars through the teaching of the late Kwang-Sae Lee, who had studied with Sellars at Yale. He also received a *Licence de philosophie* from the Université de Paris IV-Sorbonne and currently holds an assistantship at USF allowing him to play a role in a massive project undertaken by Professor Roger Ariew and others, *A New Critical Edition and Complete English Translation of the Correspondence of René Descartes,* funded by NEH and ACLS. His dissertation will explore contemporary developments of Sellarsian themes.

Michael R. Hicks, Miami University. Michael Hicks is assistant professor of philosophy at Miami University in Oxford Ohio. He earned his PhD from the Johns Hopkins University in 2009. His work in the philosophies

of language and mind has appeared in journals including *Philosophical Studies* and *Inquiry*.

Jeremy Randel Koons, Georgetown University. Jeremy Randel Koons is associate professor of philosophy at Georgetown University School of Foreign Service in Qatar. He has published widely in metaethics, epistemology, and philosophy of religion. His publications include *The Normative and the Natural* (co-authored with Michael P. Wolf; Palgrave Macmillan, 2016) and *Pragmatic Reasons: A Defense of Morality and Epistemology* (Palgrave Macmillan, 2009).

Danielle Macbeth, Haverford College. Danielle is T. Wistar Brown Professor of Philosophy at Haverford College in Pennsylvania, and the author of *Frege's Logic* (Harvard UP, 2005) and *Realizing Reason: A Narrative of Truth and Knowing* (Oxford UP, 2014). She has also published on a wide variety of issues in the history and philosophy of mathematics, philosophy of language, philosophy of mind, pragmatism, and other topics. Macbeth was a fellow at the Center for Advanced Study in the Behavioral Sciences in Palo Alto in 2002–3 and has been the recipient both of an ACLS Burkhardt Fellowship and an NEH Fellowship.

James O'Shea, University College Dublin, Ireland. James O'Shea is professor of philosophy at University College Dublin (UCD), having received his PhD at the University of North Carolina, Chapel Hill in 1992. His books include *Kant's* Critique of Pure Reason: *A Critical Guide*, ed. (Cambridge University Press, in press), *Sellars and his Legacy*, ed. (Oxford University Press, 2016), *Kant's Critique of Pure Reason: An Introduction* (Routledge, 2014), *Wilfrid Sellars: Naturalism with a Normative Turn* (Wiley/Polity, 2007), and *Self, Language, and World: Problems from Kant, Sellars, and Rosenberg*, co-edited with Eric Rubenstein (Ridgeview, 2010). He also works on Hume, pragmatism, and epistemology.

David Pereplyotchik, Kent State University. David Pereplyotchik is currently assistant professor at Kent State University. He received his BA in philosophy from New York University and his PhD in philosophy, with a concentration in cognitive science, from the City University of New York (CUNY), Graduate Center. He is the author of *Psychosyntax: The Nature of Grammar and Its Place in the Mind* (Springer, 2017), as well as several articles at the interface between philosophy and cognitive science.

Huw Price, Cambridge University, UK. Huw Price is the Bertrand Russell Professor in the Faculty of Philosophy at Cambridge University as well as a Fellow of Trinity College, Cambridge. He is author of *Facts and the Function of Truth* (Blackwell, 1998), *Time's Arrow and Archimedes' Point* (Oxford University Press, 1996), *Naturalism Without Mirrors* (Oxford University Press 2011), as well as an extensive range of journal articles. His René Descartes Lectures were published as *Expressiv-*

ism, Pragmatism and Representationalism (Cambridge University Press, 2013), with commentary essays by Simon Blackburn, Robert Brandom, Paul Horwich, and Michael Williams. He is a fellow of the British Academy and a past president of the Australasian Association of Philosophy.

Carl B. Sachs, Marymount University, Virginia. Carl B. Sachs is an assistant professor of philosophy at Marymount University. He is the author of *Intentionality and the Myths of the Given: Between Pragmatism and Phenomenology* (Routledge, 2014). He works in the intersection of phenomenology, cognitive science, and American pragmatism, with an emphasis on the Pittsburgh School. He has published on Sellars, Davidson, Rorty, McDowell, Nietzsche, Adorno, and Merleau-Ponty. He has a PhD from University of California at San Diego and is a co-founder (with Willem deVries) of the Wilfrid Sellars Society, of which he is presently vice president.

Index