

Lecture Notes in Artificial Intelligence

5466

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Flaminio Squazzoni (Ed.)

Epistemological Aspects of Computer Simulation in the Social Sciences

Second International Workshop, EPOS 2006
Brescia, Italy, October 5-6, 2006
Revised Selected and Invited Papers



Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editor

Flaminio Squazzoni
University of Brescia
Department of Social Sciences
San Faustino 74/b, 25122 Brescia, Italy
E-mail: squazzon@eco.unibs.it

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.2.11, I.2, J.4, I.6, H.4.2, K.4

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-642-01108-X Springer Berlin Heidelberg New York
ISBN-13 978-3-642-01108-5 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2009
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12638604 06/3180 5 4 3 2 1 0

Preface

This volume collects the revised versions of the invited and selected papers that were presented at the Second EPOS—Epistemological Perspectives on Simulation—Workshop, held in Brescia, Italy, in October 2006. EPOS is a bi-annual cross-disciplinary workshop on simulation originally established by Ulrich Frank and Klaus G. Troitzsch, with a first edition held in Koblenz in July 2004. EPOS aims to provide a forum for scholars from various disciplines, such as the social sciences, computer sciences, engineering and natural sciences, who are interested in discussing epistemological aspects of computer simulation across disciplinary boundaries. The common belief behind the workshop is the recognition that the time has come to seriously reflect on epistemological and methodological preconditions, processes and consequences of simulation as a research tool.

During the first edition in Koblenz 2004, a number of interesting topics were carefully addressed: the link between theory and simulation models, the empirical validation of agent-based models in the natural and the social sciences, the relation between models and truth, as well as the role of stylized facts in evidence-based models. A good cross-disciplinary atmosphere permeated the workshop, making possible the exchange of knowledge and ideas beyond any disciplinary boundary. The first EPOS proceedings were edited by Ulrich Frank and Klaus G. Troitzsch and published in the *Journal of Artificial Societies and Social Simulation*, Vol. 8, No. 4, 2005.

The second edition in Brescia 2006 was led by social scientists, after a careful and strong selection process that limited the number of presented papers from 35 submissions to 11 presentations. Nigel Gilbert and Rosaria Conte were the invited speakers. Topics addressed ranged from epistemological and methodological contents, such as the relevance of empirical foundations for agent-based simulations, the role of theory, the concepts and meanings of emergence, the trade-off between simplification and complexification of models. The discussion among the participants was vivid and vibrant, confirming the common interest in these issues. Given the high-level contents of the discussions, together with Ulrich Frank and Klaus G. Troitzsch, who attended the workshop, we decided to write an introductory chapter, where most of the contents of the discussions have a representative synthesis.

November 2008

Flaminio Squazzoni

Organization

The Second EPOS—Epistemological Perspectives on Simulation—Workshop was held at the University of Brescia in October 2006 with the financial support of the Lucchini Foundation, University of Brescia and ESSA (European Social Simulation Association), for which we gratefully acknowledge Severo Bocchio (Lucchini Foundation Chief), Giancarlo Provasi (Vice-Chancellor of the University of Brescia), Serafino Negrelli (University of Brescia), and the members of ESSA Management Committee.

Scientific Chair

Flaminio Squazzoni

Program Committee

Petra Ahrweiler
Frédéric Amblard
Riccardo Boero
François Bousquet
Thomas Brenner
Edmund Chattoe-Brown
Rosaria Conte
Alexis Drogoul
Bruce Edmonds
Andreas Flache
Ulrich Frank

Nigel Gilbert
László Gulyás
David Hales
Scott Moss
Mario Paolucci
Béla Pátkai
Andreas Pyka
Nicole Saam
Takao Terano
Pietro Terna
Klaus G. Troitzsch

Additional Referees

Kai-H Brassel
Marco Castellani
Guido Fioretti

Rainer Hedgselmann
Matteo Richiardi
Juliette Rouchier

Organization Staff

Marco Castellani
Beatrice Marelli
Mimma Negretti

Domenico Panetta
Marisa Scaringi
Viviana Zenucchini

Table of Contents

EPOS-Epistemological Perspectives on Simulation: An Introduction	1
<i>Ulrich Frank, Flaminio Squazzoni, and Klaus G. Troitzsch</i>	
Invited Papers	
The Epistemologies of Social Simulation Research	12
<i>Nigel Gilbert and Petra Ahrweiler</i>	
From Simulation to Theory (and Backward)	29
<i>Rosaria Conte</i>	
Selected Papers	
Talking about ABSS: Functional Descriptions of Models	48
<i>Scott Moss</i>	
What Does Emergence in Computer Simulations? Simulation between Epistemological and Ontological Emergence	60
<i>Alex Schmid</i>	
Emergence as an Explanatory Principle in Artificial Societies. Reflection on the Bottom-Up Approach to Social Theory	69
<i>Martin Neumann</i>	
Reconstruction Failures: Questioning Level Design	89
<i>Camille Roth</i>	
Narrative Scenarios, Mediating Formalisms, and the Agent-Based Simulation of Land Use Change	99
<i>Nicholas M. Gotts and J. Gary Polhill</i>	
Validation and Verification in Social Simulation: Patterns and Clarification of Terminology	117
<i>Nuno David</i>	
Validation and Verification of Agent-Based Models in the Social Sciences	130
<i>Paul Ormerod and Bridget Rosewell</i>	
Abductive Fallacies with Agent-Based Modeling and System Dynamics	141
<i>Tobias Lorenz</i>	

Algorithmic Analysis of Production Systems Used as Agent-Based Social Simulation Models	153
<i>Jim Doran</i>	
The Nature of Noise	169
<i>Bruce Edmonds</i>	
Author Index	183

EPOS-Epistemological Perspectives on Simulation: An Introduction

Ulrich Frank¹, Flaminio Squazzoni², and Klaus G. Troitzsch³

¹ University of Duisburg-Essen, Germany
ulrich.frank@uni-due.de

² Department of Social Sciences, University of Brescia
Via San Faustino 74/B, 25122 Brescia, Italy
squazzon@eco.unibs.it

³ University of Koblenz-Landau, Germany
kgt@uni-koblenz.de

1 Introduction

There is strong evidence that computer simulation is increasingly recognized as an important analytical tool in many social sciences disciplines and fields. During the last ten years, a number of new journals, which are devoted to this field, have been founded and others have increased their influence (i.e., JASSS, CMOT, Social Science Computer Review, Autonomous Agent and Multi-Agent Systems, Journal of Economic Dynamics and Control, Computational Economics, Computational Management Science). Special issues and extensive reviews of the literature have been published in influential and standard journals¹. At the same time, new international associations and societies were born, with an increasing number of members (i.e., ESSA in Europe, NAACSOS in North America, The Society for Computational Economics), many research centers and institutes have been successfully launched², many workshops, conferences and congresses are organized every year (with the first world congress on social simulation held in 2006 in Tokyo and the second one in Washington in 2008), and an open market of tools and simulation platforms (i.e., Swarm, Repast, Ascape, NetLogo), based on a vast community of developers and users, is steadily growing.

¹ Special issues in «American Behavioral Science» in 1999, «Journal of Economic Dynamics and Control» 2001 and 2004, «Computational Economics» 2001 and 2007, «Proceedings of the National Academy of Sciences» 2002, «Journal of Economic Behavior and Organization» 2004, «Journal of Public Economic Theory» 2004, «American Journal of Sociology» 2005; «Advances on Complex Systems» 2008; reviews and/or papers in «Nature», «Science», «Physica A.», «Journal of Theoretical Biology», «American Sociological Review», «Annual Review of Sociology», «Philosophy of the Social Sciences», «Artificial Intelligence Review».

² For example: Santa Fe Institute in New Mexico, USA; Center on Social and Economic Dynamics, at the Brookings Institution, Washington, USA; Centre for Policy Modeling in Manchester, UK; Centre for Research in Social Simulation in Guildford, UK; Laboratory of Agent-Based Social Simulation, CNR, Rome, Italy; Center for Social Complexity at George Mason University, Virginia, USA; Center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon, Pittsburgh, USA; Center for the Study of Complex Systems, University of Michigan, USA.

In these last years, a significant number of influential papers, handbooks, tutorials and books on simulation have been published in many disciplines. For instance, in the social sciences, Axelrod (1997) and Nowak and Sigmund (1998; 2005), independent of each other, launched a very influential research program on the study of cooperation and reciprocity via computer simulation, with numerous followers not only in the social sciences. Epstein (2007) recently suggested a coherent manifesto to what he called “generative social science”, where agent-based models are used as the best-suited modeling instrument to formalize micro-macro models of social and economic phenomena. In his influential manifesto of analytical sociology, Hedström (2005) did the same, locating agent-based models at the core of his programmatic argument. Sawyer (2005) argued the relevance of computer simulation to understanding social emergence. From a methodological perspective, de Marchi (2005) suggested a coherent approach to integrate computational methods with statistical and empirical methods in the social sciences.

Agent-based computational economics is a well recognized field, especially by evolutionary, institutional and applied economists (Saviotti 2003; Pyka and Arhweiler 2004), with good examples of relevant handbooks, tutorials or methodological books on computer simulation and modeling (i.e., Kendrick, Mercado and Amman 2005; Tesfatsion and Judd 2006). In geography and urban studies, there is currently an effective cross-fertilization between computer simulations and other methods and tools, like Geographical Information Systems and statistical surveys and data (i.e. Gimblett 2002; Batty 2005). Computer simulation is used also in anthropology, ecology and in environmental studies (Janssen 2002; Volker and Railsback 2005). In organizational and management studies (Prietula, Carley and Gasser 1998; Lomi and Larsen 2000), as well as in business engineering and industrial processes planning and control (i.e. van Elst, Abeker and Dignum 2004), computer simulation is a well-established tool and an innovative method. Scholars in the areas of artificial intelligence and robotics are using simulation to explore and design intelligent cooperative systems (i.e., Bonabeau, Dorigo and Theraulaz 1999). A good number of general textbooks, tutorials and introductory books on the usage of computational tools and modeling have been recently published, contributing to the promotion of such an approach in the new generation of scholars (i.e., Suleiman, Troitzsch and Gilbert 2000; Wooldridge 2002; Gilbert and Troitzsch 2005; Shiflet and Shiflet 2006; Gilbert 2008).

This introduction aims to present an overview on epistemological debates on computer simulation in the social sciences. Does computer simulation mean a new scientific paradigm or is it just a tool? What is the difference between simulation modeling and traditional analytic models? Does simulation allow for a new relation between theory and data? What are the preconditions for a proper use of simulation? These are just a few questions on which the community is recently focusing (Frank and Troitzsch 2005).

In this introduction, we argue that simulation constitutes an inspiring new research perspective in the social sciences. As we see in the following sections, as well as in most chapters of this book, there are sound substantive and methodological reasons for this argument. First, the possibility of using the computer as a tool for formalizing generative models of social phenomena, with the aim to foster theory building, exploration and testing, allows us to enlarge the space of application of analytical approaches in the social sciences. Simulation indeed requires formal models of social

phenomena, thereby contributing to a rationalistic approach to research. Thus it provides an alternative to those streams in the social sciences that are dominated by post-modernism and by all those approaches that deny the possibility of using formalized models in the social sciences. Secondly, by deploying computers, simulation offers the chance to produce - and test - patterns of knowledge that would hardly be possible without this tool. Thirdly, agent-based simulation is a way to cope with complexity and to develop rich, differentiated scenarios at the same time, with several positive consequences at the level of realism and empirical foundation of models.

To give a clear picture of this innovation, the second section summarizes a set of epistemological and methodological open issues on which the community is carefully focusing at the present. The debate is mostly expressed in terms of vivid disputes on dichotomies as follows: generative vs. causal explanations via simulation; meanings and forms of emergence, with a confrontation of ontological and epistemological meanings; theoretical abstraction vs. empirical foundations; theoretical vs. practical purposes. As we see, most of these issues have been explicitly addressed by the authors of this book, making possible to undertake some step forward. As a matter of fact, computer simulation helps secularize such philosophical debates by bringing them away from a mere foundational and philosophical to a more pragmatic and evidence-oriented level. The varied, even contradictory, positions are an indication of this burgeoning, innovative field of research.

In the third section, we propose our assessment of the future of this field of research. In particular, according to the relevant issues of the present, we attempt to identify some epistemological, methodological and substantive puzzles which need to be solved. It is our firm belief that the solution to these puzzles would improve the scientific progress of simulation. Finally, the fourth section offers a description of the book chapters, which are all devoted to clarifying relevant epistemological and methodological issues.

2 Epistemological and Methodological Issues

Computer simulations, and specifically agent-based models, should be viewed as an epistemological innovation, because they permit and promote a generative approach to the modeling and the explanation of social phenomena, strengthening an agent-based paradigm to social science (Epstein and Axtell 1996; Epstein 2007). A “generative approach” is based on the following constituents:

1. The target of the model is a macro phenomenon of interest, let us say a given statistical empirical regularity, evidence or a stylized fact;
2. The model comprehends a set of theoretical hypotheses about a population of agents, rules, constraints and interaction structures, which the modeler believes to be a system that maps those generative causal processes that are supposed to be responsible for the phenomenon;
3. The simulation of the model allows us to produce a set of artificial data, which are controlled by means of a set of indicators, and which are also used to adjust and verify the initial hypotheses;
4. These data are then compared to the empirical data, evidence, or to stylized facts that are the explanatory target of the modeler.

The epistemological assumption is if we are able to generate or grow a macro phenomenon via an agent-based model, on the basis of theoretically or empirically plausible assumptions on the micro-foundations, then we can consider these assumptions as sufficient, even if not necessary conditions for explaining it. Therefore, the simulation model is seen as an engine to find *candidate* explanations of macro social phenomena (Epstein 2007). This position is a kind of standard in social simulation.

2.1 Generative Explanation

As the readers will see by going through some chapters of this book, the generative assumption has been both criticized and developed on two main facets: the difference between a generative and a causal explanation and the difference between a “pure” theoretical explanation and a multi-level empirically-founded model.

Regarding the first facet, it is standard practice to assume that a generative explanation would differ from a standard covering law, a statistical or a narrative explanation because the former would be based on a mechanisms-based explanation (Goldthorpe 2000; George and Bennett 2004). In this perspective, explaining a macro empirical phenomenon of interest means referring to a set of entities and processes (agents, action and interaction) that are spatially and temporally organised in such a way that they regularly bring about that phenomenon (Hedström and Swedberg 1998). In other words, a social mechanism is “a constellation of entities and activities that are linked to one another in such a way that they regularly bring about a particular type of outcome” (Hedström 2005, 11). As George and Bennett correctly pointed out (2004, 141), the difference between a law and a mechanism means a difference between a static correlation (“If X, then Y”) and a specification process (“X leads to Y through steps A, B, C”) (George and Bennett 2004, 141).

In order to take such a process into account in the social sciences, it is necessary to identify what James Coleman (1990) defined as the chains of macro-micro-macro links: macro-micro situation mechanisms, micro-micro interaction mechanisms, and micro-macro transformation mechanism. According to the so called “Coleman bath tub”, a mechanisms-based explanation should first dissect and then combine the jointed action of situational mechanisms (macro-micro), action-formation mechanisms (micro-micro), and transformational mechanisms (micro-macro). The analogy of the bath tub symbolizes that macro outcomes are what can be seen on surface, hopefully expressed in formal terms of empirical regularities, such as statistical properties or stylised facts. But, to understand why and how the surface moves, researchers should go under the surface, to see the causal engine of the movement (Squazzoni 2008).

It is well known that mechanisms are often brought about by bundles or configurations of mechanisms, “some of which contribute to the effect and some of which may operate to counteract the effect or reduce its magnitude”. The following is a simple example, which has been reported by Paul Humphreys:

“a car went off a road at a curve because of excessive speed and the presence of sand on the road and despite clear visibility and an alert driver. He notes that the addition of another mechanism or contextual factor can change a contributing cause to a counter-acting one, or vice-versa: sand decreases traction on a dry road, but increases traction when there is ice on the road” (George and Bennett 2004, 145-146).

If this is true, how is it possible to specify what kind of mechanism is working in the phenomenon of interest, and in what direction they are conducting the bath tub? There is no space to get into more detail on this point, but the argument we would like to emphasize here is as follows: a) the possibility to use an agent-based model to generate a phenomenon does not automatically mean to be in a position to have a causal explanation; b) given the famous arguments of the so-called “multiple realizability” (George and Bennett 2004; Sawyer 2005), according to which a mechanisms-based explanation closely depends on the capability of combining theoretical assumptions and empirical specifications. This is based on the famous arguments of the so-called “multiple realizability” (George and Bennett 2004; Sawyer 2005): the same micro-mechanism could produce different macro outcomes, and the same macro outcome could be produced by different micro mechanisms, Just the reference to empirical evidence can help discriminate among potential alternative theoretical explanations.

2.2 Empirical Foundations and Validation of Simulation Models

There is growing literature on the quest of the empirical foundations and validation of computer simulation models, in particular in the field of agent-based models (i.e., Boero and Squazzoni 2005; Windrum, Fagiolo and Moneta 2007). As the reader will see, four of the chapters of this book deal exactly with this topic. The question is: is the theory behind a model really enough to do good science?

Social simulation researchers seem implicitly split on this question. One faction considers computer simulation as a method for theory building and exploration, as though the computer is a complete substitute for laboratory experiments and empirical analyses. The other faction denies that computer simulation is a self-exhaustive experimental tool and believe that the purpose of social science, also that of computational social science, is primarily to explain empirical phenomena, with the recourse to empirical data.

The first faction believes in the intrinsic added value of the “artificial societies” as a means to strengthen the theoretical foundations of social sciences. This faction believes that the problem with the development of the social sciences lies in the weakness of the theoretical pillars. This faction believes in the “synthetic” features of simulation, i.e. in the use of computer simulation to synthesize all the aspects of the complex nature of social life. The second faction believes that computer simulation is just a tool that allows the social scientist to explain empirical phenomena, a tool that should be used with other tools and empirical methods. This faction believes that the problem lies in the excess of theoretical accounts and models, and the need for strengthening the capacity of modeling and understanding specific empirical phenomena.

This is basically the epistemological side of the debate on the empirical foundations and validation of simulation models. Some chapters of this book explicitly or implicitly focus on this side (Chapter 2 by Rosaria Conte, Chapter 4 by Scott Moss and Chapter 10 by Paul Ormerod and Bridget Rosewell). Regarding the methodological side, a good number of contributions and examples have recently started to elucidate part of the complicated implications of available methodological practices (i.e., Moss and Edmonds 2005; de Marchi 2005; Windrum, Fagiolo and Moneta 2007). For instance, given the micro-macro link that dominates the social phenomena, an interesting methodological issue is the need for a multi-level validation of the model,

where empirical data on the macro level are combined with empirical data on the micro level (Boero and Squazzoni 2005). The challenge is to understand how to foster standard methodological practices of validation that could be shared and tested by the community.

2.3 Emergence

The concept of emergence is tied up with the concept of simulation (Sawyer 2005). Two of the chapters in this book try to explicitly deepen such a tie (Chapters 5 and 6 by Alex Schmid and Martin Neumann). In fact, emergence is closely related to “properties”, “structures”, “regularities”, “patterns” and “phenomena” that are macro consequences of basic generative processes that can only be explored via simulations (Bedau 1997). Therefore, it seems that simulation is a suitable approach to reconstruct and explain emergence.

From an epistemological viewpoint, the debate in social simulation seems to be stuck in the middle of two extremes (Squazzoni 2008): the ontological meaning of emergence and the micro-reductionism.

A first faction of social scientists vindicates the reductionistic approach to simulation, refusing any possible ontological meaning of emergence (Hedström 2005; Epstein 2007). The position is extremely coherent: if a theoretical explanation of a macro social phenomenon of interest means the creation of a simulation model to explore plausible theoretical micro assumptions that generate and explain the phenomenon, then macro patterns, phenomena, or processes under study are the resultant implications of micro-assumptions explored inside a formalised model. Particularly, counter-intuitive and intriguing macro phenomena, which would be said “to emerge” by the supporters of social emergence, can be in principle and *de facto* reduced to a set of mechanisms at the lower level (Kim 2006). In the social sciences, such a position totally conforms to the methodological individualism (Coleman 1990), where the ontological existence of “invisibles” or “social entities” and their supposed “causal power” are ontologically and methodologically rejected (Hedström 2005). It is recognized that social science theory requires the explicit generative reference to individual action and interaction to avoid the risk of reifying and evoking macro categorical concepts, such as “class”, “culture”, or “system”, which are often used as explanatory categories in a tautological way. The division between the micro and the macro levels of a model and the analysis of their relation are simply epistemological practices executed by the scientist for analytic purposes, and not an ontological feature of the social reality in itself (Gilbert 2005; Squazzoni 2008).

On the contrary, a second faction of scientists suggests a kind of radical emergentism (Sawyer 2005). Strongly based on a long tradition in philosophy, as well as on classic macro sociology, this faction argues the ontological nature of social emergence. Societal phenomena would belong to a different ontological level of reality compared to individual phenomena. This results in social scientists replacing the causal determinism of the micro level of the reductionists with the causal determinism of the macro level, assigning macro causal power and ontological autonomy to social entities.

In the middle, there is the pragmatic attitude of many social simulation researchers who see emergent properties and phenomena simply as epistemological features of

the scientific inquiry. This last faction refuses to accept the ontological meaning of emergence, the supposed causal power of macro social entities, and even the idea that emergence properties are intrinsic features of the reality (Bedau 1997). They instead believe that emergence is a concept with an epistemological value in itself and is created by the scientist with the aim to identify and refer to macro dynamics and change in quantitative and qualitative terms (Gilbert 2002; 2005). Social simulation strengthens links and integrative frameworks, and “secularises” this dispute, by bringing it away from a foundational and philosophical level to a more pragmatic one (Squazzoni 2008). In any case, as some of the chapters of this book demonstrate (Chapter 3 by Rosaria Conte and the three chapters on emergence by Alex Schmid, Martin Neumann and Camille Roth), the dispute is still wide open and deserves further contributions.

3 Future Prospects

The relation between theory, models and empirical validation should be considered as the next frontier for the advancement of social simulation. If theoretical abstraction should be naturally viewed as an essential part of scientific research, the development of computer simulations in the social sciences should be related to their capacity to foster the explanation of empirical phenomena. This is to avoid the risk of the self-referential nature of theoretical models as well.

If this is true, a challenge for the future is the appropriate combination of computer simulation with other methods. Instead of seeing computer simulation as a different method, or as an exhaustive method *per se*, a good practice would be to work towards combining different formal, experimental, statistical and empirical methods and try to exploit the best of each one (de Marchi 2005). The epistemological consequence would be also to expand upon the analysis of the qualities and properties that perhaps make computer simulation different from other methods and tools. But the recognition of these prospective differences does not imply, as thought by many computational social scientists, that computer simulation models should be necessarily judged outside the standard framework and practices of science.

From a methodological viewpoint, many scholars have already recognized a future prospect in the field of verification/ validation/ sharing/ communication of models and results (Axelrod 2006; Windrum, Fagiolo and Moneta 2007). As Gintis correctly outlined (2007), computer simulation models, in particular agent-based models, are profoundly penalized within the standard scientific community because of many problems of scientific communication and inter-subjective control. The case of economics is emblematic. In economics, agent-based models are sometimes introduced as theoretical models to be used as an alternative to formal analytical economic theory, sometimes as empirical models from which one could infer properties that could be analytically systematized, or on which one could build comparisons among particular cases. In any case, the lack of transparency and the impossibility/difficulty to enter in all the details of a model and of a set of simulations make evaluation of a simulation model often impossible. Only an expert in the computer language used or the expert who designed precisely those simulations presented in that paper understand the details. The result may be that referees of journals are forced to take the author’s

assertions on faith alone. This is where standard analytic models are superior to agent-based computational ones at the present. Moreover, let us suppose that a model is the result of a collaboration between a social scientist and a computer scientist, as often happens. In such a case, the problem of the verification is generally underestimated. Nuno David correctly emphasizes this in his chapter: there is a risk of underestimating possible misunderstandings and hidden differences between the theoretical and the implemented model, which are a natural source of theoretical invalidation of simulation results and, as a consequence, of the theoretical inferences conducted on them.

This observation strongly supports our final claim that the future of social simulation and the appropriate usage of computer simulation in the social sciences will depend on the investment that the leading scholars and institutions in this field will do on the training and education of young scholars, PhD and undergraduate students in simulation languages. This is necessary for raising a generation of young social scientists able to advance the scientific frontier of social simulation and to embed computer simulation in the standard toolkit of the social scientist of the 21st century. What is needed is to slowly but surely get out of the current “hand-crafted” phase, to be able to enter in a new phase, where standard practices, methods and scientific communication can get stronger and cumulate step-by-step.

4 Summary of the Book

The chapters of this book are very heterogeneous in their contents, but they can all be subsumed under the same general purpose. They are good examples of the present quest for the pre-conditions, processes and consequences of computer simulation in the social sciences, combined with the methodological, epistemological and substantive issues.

The second chapter, “The Epistemologies of the Social Simulation Research” is an invited contribution by Nigel Gilbert and Petra Ahrweiler. The authors aim to fertilize the epistemological debate on social simulation with the philosophical debate in the social sciences. They conclude with a critical and pragmatic position on the need for an epistemological and methodological unified perspective, arguing that evaluation criteria for simulation research need to be tuned to research aims, methodological positions and domains of study, rather than being conditioned by epistemological and/or methodological pre-established viewpoints.

The third chapter, “From Simulation to Theory (and Backward)” is an invited contribution by Rosaria Conte. It focusses on the “generative approach” to computational social sciences, which has been synthesized in the recent “generativist manifesto” by Epstein (2007). The author argues against the tendency to conflate theory and simulation, as well as explanation and computational generativeness. She proposes a body of arguments to emphasize the relevance of a computational social science theory that should be able to combine bottom-up theory of general local rules and top-down downward causation. In doing so, she locates herself in the recent debate on the distinctive features of social emergence (Sawyer 2005).

The fourth chapter, “Talking about ABSS: Functional Descriptions of Models”, by Scott Moss, attempts to integrate empirical and abstracted models, with the aim of reducing the divide between the practices and the approaches of empirical and theoretical

analyses in social simulation. The author suggests a framework to foster the discussion of epistemological issues, such as prediction, validation and verification.

The fifth, sixth and seventh chapters constitute a sub-section focused on emergence. Alex Schmid proposes a review of ontological and epistemological meanings of emergence, with a contextualization of the particular role of computer simulation as an analytical tool for studying emergent properties and processes, while Martin Neumann reassesses the literature to propose a conceptual framework for understanding the autonomy of the emergent properties in social spheres. Finally, Camille Roth argues that reductionism and emergentism fail at the same time in detecting ill-conceived modeling ontology.

The eighth chapter, “Narrative Scenarios, Mediating Formalisms, and the Agent-Based Simulation of Land Use Change” by Nick Gotts and J. Gary Polhill, inaugurates the methodological chapters of this book. The authors move on from the evidence that the kinds of systems studied through agent-based models in the social sciences are first understood in terms of narrative languages and descriptions, with the consequence that a set of problems arise because of the difference between the semantics of natural language and programming languages. Using examples taken from simulation of land use change, they articulate an ontological framework and a formalism to mediate between the two. The following two chapters focus on the relevant quest of the verification/validation of simulation models. Nuno David suggests distinguishing between verification and validation, as well as clarifying their meanings and their relation. Moreover, he tries to embed the methodological debate on the relevant quest of what “truth” and what “knowledge” are generated by means of simulations. Paul Ormerod and Bridget Rosewell cope with the quest of the empirical validation of agent-based models in the social sciences. They reconstruct some of the difficulties in establishing verification and validation of agent-based models, and suggest simplifying the theoretical building blocks of the models at most, so that comparison and acceptance of models could be improved. Moreover, they bring empirical evidence that supports the so called KISS (Keep It Simple, Stupid) principle of agent-based modeling, namely that agents in real situations tend to act intuitively rather than rationally. In this perspective, they criticize the supposed “naturalness” of the cognitive approach to agent-based models.

The eleventh chapter, “Abductive Fallacies with Agent-Based Models and System Dynamics” by Tobias Lorenz, elaborates on the abductive nature of simulation, with particular attention to the role of simulation methodologies in conditioning the modeling process. The twelfth chapter, “Algorithmic Analysis of Production Systems Used as Agent-Based Social Simulation Models” by Jim Doran, is an exploration of the algorithmic analysis of simulation models as a means to systematize model abstraction. To demonstrate the potential of his exploration, the author refers to an applied example, namely the Iruba model of a guerrilla war. Last but not least, the thirteenth chapter, “The Nature of Noise”, by Bruce Edmonds, works up a comprehensive exploration on the concept of noise. The author argues that noise and randomness are *not* the same and thoughtlessly conflating them can result in misleading assumptions. This exploration has relevant implication for the modeling and understanding of complex phenomena.

Acknowledgments. This introduction owes much to Alexandra Bohnet (University of Koblenz-Landau) who improved grammar and style of the text.

References

1. Axelrod, R.: *The Complexity of Cooperation. Agent-Based Models of Competition and Collaboration*. Princeton University Press, Princeton (1997)
2. Axelrod, R.: *Advancing the Art of Simulation in the Social Sciences*. In: Rennard, J.P. (ed.) *Handbook of Research on Nature Inspired Computing for Economy and Management*. Idea Group, Hersey (2006)
3. Batty, M.: *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. MIT Press, London (2005)
4. Bedau, M.: *Weak Emergence*. *Philosophical Perspectives* 11, 375–399 (1997)
5. Boero, R., Squazzoni, F.: *Does Empirical Embeddedness Matter? Methodological Issues on Agent-Based Models for Analytical Social Science*. *Journal of Artificial Societies and Social Simulation* 8(4) (2005), <http://jasss.soc.surrey.ac.uk/8/4/6.html>
6. Bonabeau, E., Dorigo, M., Theraulaz, G.: *Swarm Intelligence From Natural to Artificial Systems*. Santa Fe Institute Studies in the Sciences of Complexity. Oxford University Press, Oxford (1999)
7. Coleman, J.: *Foundations of Social Theory*. Cambridge University Press, Cambridge (1990)
8. de Marchi, S.: *Computational and Mathematical Modeling in the Social Sciences*. Cambridge University Press, Cambridge (2005)
9. Epstein, J.M.: *Generative Social Science. Studies in Agent-Based Computational Modeling*. Princeton University Press, Princeton (2007)
10. Epstein, J.M., Axtell, R.: *Growing Artificial Societies. Social Science from the Bottom-Up*. MIT Press, Cambridge (1996)
11. Frank, U., Troitzsch, K.G.: *Epistemological Perspectives on Simulation*. Special Issue, *Journal of Artificial Societies and Social Simulation* 8(4) (2005), <http://jasss.soc.surrey.ac.uk/8/4/7.html>
12. George, A.L., Bennett, A.: *Case-Studies and Theory Development in the Social Sciences*. MIT Press, Cambridge (2004)
13. Gilbert, N.: *Varieties of Emergence*. In: Sallach, D. (ed.) *Social Agents: Ecology, Exchange, and Evolution*. Agent 2002 Conference, University of Chicago and Argonne National Laboratory, pp. 41–56 (2002)
14. Gilbert, N.: *When Does Social Simulation Need Cognitive Models?* In: Sun, R. (ed.) *Cognition and multi-agent interaction: From cognitive modeling to social simulation*. Cambridge University Press, Cambridge (2005)
15. Gilbert, N., Troitzsch, K.G.: *Simulation for the Social Scientists*, 2nd edn. Open University Press, Buckingham Philadelphia (2005)
16. Gilbert, N.: *Agent-Based Models*. Sage Publications, London (2008)
17. Gimblett, R.H.: *Integrating Geographic Information Systems and Agent-based Modeling Techniques for Simulating Social and Ecological Processes*. Oxford University Press, Oxford (2002)
18. Gintis, H.: *Review of Handbook of Computational Economics, Volume II: Agent-Based Computational Economics by Tesfatsion L. and Judd K.J.* *Journal of Artificial Societies and Social Simulation* 10(1) (2007), <http://jasss.soc.surrey.ac.uk/10/1/reviews/gintis.html>

19. Goldthorpe, J.H.: *On Sociology. Numbers, Narratives and the Integration of Research and Theory*. Oxford University Press, Oxford (2000)
20. Hedström, P.: *Dissecting the Social. On the Principles of Analytical Sociology*. Cambridge University Press, Cambridge (2005)
21. Hedström, P., Swedberg, R. (eds.): *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge University Press, Cambridge (1998)
22. Kendrick, D.A., Mercado, P.R., Amman, H.M.: *Computational Economics*. Princeton University Press, Princeton (2005)
23. Kim, J.: *Emergence: Core Ideas and Issues*. *Synthese* 151, 547–559 (2006)
24. Janssen, M.A. (ed.): *Complexity and Ecosystem Management: the Theory and Practice of Multi-Agent Systems*. Edward Elgar Publishing, Cheltenham (2002)
25. Lomi, A., Larsen, E.R. (eds.): *Dynamics of Organizations: Computational Modeling and Organization Theories*. MIT Press, Cambridge (2000)
26. Moss, S., Edmonds, B.: *Sociology and Simulation: Statistical and Qualitative Cross-Validation*. *American Journal of Sociology* 110(4), 1095–1131 (2005)
27. Nowak, M.A., Sigmund, K.: *The Dynamics of Indirect Reciprocity*. *Journal of Theoretical Biology* 194, 561–574 (1998)
28. Nowak, M.A., Sigmund, K.: *Evolution of Indirect Reciprocity*. *Nature* 437, 1291–1298 (2005)
29. Prietula, M., Carley, K., Gasser, L. (eds.): *Simulating Organizations: Computational Models of Institutions and Groups*. MIT Press, Cambridge (1998)
30. Pyka, A., Arhweiler, P.: *Applied Evolutionary Economics and Social Simulation*. *Journal of Artificial Societies and Social Simulation* 7(2) (2004), <http://jasss.soc.surrey.ac.uk/7/2/6.html>
31. Saviotti, P.P. (ed.): *Applied Evolutionary Economics: New Empirical Methods and Simulation Techniques*. Edward Elgar Publishing, Cheltenham (2003)
32. Sawyer, R.K.: *Social Emergence. Societies as Complex Systems*. Cambridge University Press, Cambridge (2005)
33. Shiflet, A.B., Shiflet, G.W.: *Introduction to Computational Science: Modeling and Simulation for the Sciences*. Princeton University Press, Princeton (2006)
34. Squazzoni, F.: *The Micro-Macro Link in Social Simulation*. *Sociologica*, 1, 2 (2008), <http://www.sociologica.mulino.it/journal/article/index/Article/Journal:ARTICLE:179>
35. Suleiman, R., Troitzsch, K.G., Gilbert, N. (eds.): *Tools and Techniques for Social Science Simulation*. Springer, Heidelberg (2000)
36. Tesfatsion, L., Judd, K.L. (eds.): *Handbook of Computational Economics, Volume II: Agent-Based Computational Economics*. Elsevier/North-Holland, Amsterdam (2006)
37. van Elst, L., Dignum, V., Abecker, A. (eds.): *AMKM 2003. LNCS, vol. 2926*. Springer, Heidelberg (2004)
38. Volker, G., Railsback, S.F.: *Individual-Based Modelling and Ecology*. Princeton University Press, Princeton (2005)
39. Windrum, P., Fagiolo, G., Moneta, A.: *Empirical Validation of Agent-Based Models: Alternatives and Prospects*. *Journal of Artificial Societies and Social Simulation* 8(4) (2007), <http://jasss.soc.surrey.ac.uk/10/2/8.html>
40. Wooldridge, M.: *An Introduction to Multi-Agent Systems*. John Wiley and Sons Limited, Chichester (2002)

The Epistemologies of Social Simulation Research*

Nigel Gilbert¹ and Petra Ahrweiler²

¹ CRESS - Centre for Research in Social Simulation, University of Surrey, UK
n.gilbert@surrey.ac.uk

² Innovation Policy Unit, National Institute of Technology Management,
University College Dublin, Ireland
petra.ahrweiler@ucd.ie

Abstract. What is the best method for doing simulation research? This has been the basis of a continuing debate within the social science research community. Resolving it is important if the community is to demonstrate clearly that simulation is an effective method for research in the social sciences. In this paper, we tackle the question from a historical and philosophical perspective. We argue that the debate within social simulation has many connections with the debates that have echoed down the years within the wider social science community about the character of social science knowledge and the appropriate epistemological and methodological assumptions on which social science research should rest.

Keywords: Social simulation, epistemology, *methodenstreit*, Max Weber.

1 Introduction

What is the best method for doing simulation research? This has been the basis of a continuing debate within the social science research community. Resolving it is important if the community is to demonstrate clearly that simulation is an effective method for research in the social sciences. In this paper, we tackle the question from a historical and philosophical perspective. We argue that the debate within social simulation has many connections with the debates that have echoed down the years within the wider social science community about the character of social science knowledge and the appropriate epistemological and methodological assumptions on which social science research should rest.

In section 1, we review the debate that was conducted primarily within German political economy at the beginning of the twentieth century about the degree to which it is appropriate to search for ‘laws’ of society. In section 2, we show that many of the

* An earlier version of this chapter was presented under title, ‘Kiss and tell: in praise of abstraction? Starbucks or Café Florian’ to the workshop on Epistemological Perspectives on Simulation - II, held in Brescia, Italy on October 25-26, 2006. We are indebted to the participants at the workshop, and especially Flaminio Squazzoni, for their good company, their wise comments and their hospitality. This chapter is a sequel to a paper, ‘Caffè Nero: the Evaluation of Social Simulation’, presented at the first EPOS workshop and subsequently published as Ahrweiler and Gilbert (2005).

arguments formulated then by luminaries such as Weber, Schumpeter, Schmoller and Menger are still relevant to the current debate about methodological approaches to social simulation. In section 3, we propose that these arguments lead to the conclusion that, *contra* to some present-day commentators, there is no one best epistemological or methodological approach. We conclude that evaluation criteria for simulation research need to be tuned to research aims, methodological positions and domains of study, and it is not right to judge all simulation research using the same set of evaluation instruments.

2 The Knowledge Claims of the Social Sciences: Mathematics or History

2.1 Background: Rationalism and Empiricism

How can we be sure about our knowledge of the world? Two contrasting answers have been proposed, under the names of rationalism and empiricism. Rationalism says that it is reason or rationality that provides certainty. Reason is constituted by two properties: the first is an intuition concerning final knowledge foundations, that is, the ultimate principles and ideas that form a set of axiomatic-terminological conditions for all further knowledge and theory building; the second is the human ability to engage in deductive reasoning. Starting from axiomatic principles, we can extend our knowledge to yield theoretical systems (Specht in Schnaedelbach 1984: 72ff) such as mathematics, the discipline that for Descartes represented the idea of science (gr. *mathema* = science).

In contrast, the common denominator of empiricism is reference to the senses: certainty of knowledge can only be achieved by the direct experience of an individual case. Locke challenged Descartes with “*Nihil est in intellectu, quod non prius fuerit in sensu*” (Nothing is in the understanding that was not earlier in the senses). However, the outcome of inductive reasoning is problematic: each generalising statement that goes beyond a limited class of individual cases is uncertain and its truth completely depends on its empirical foundation. For Hume’s radical empiricism, inductive reasoning was therefore out of the question. The same scepticism can later be found in Popper’s “We do not know, we just guess”. In his view, theories in empirical sciences cannot be inductively verified but – using deductive principles – they can be contradicted by elementary observations, that is, they can be falsified.

The debate between rationalists and empiricists over the foundations of knowledge is still unsolved. In the following, we enter a version of this debate at a time that was important for sociology and economics as they became established as academic disciplines.

2.2 The Methodological Dispute (Methodenstreit) in Continental Political Economy

In the 19th century, the French philosopher Auguste Comte (1798-1857) had postulated that social philosophy should leave all speculative ideas and follow the positivist example of natural science: “for positivist philosophy all processes follow deterministic laws.

There is no use in looking for first causes or final purposes” (Comte 1974: 5). To investigate the laws of the social realm, Comte introduced the notion of *Sociologie* which was interpreted by him as “social physics”. In political economy in the 1850s, the so-called *elder historical school* (Wilhelm Roscher, Karl Knies) had tried to establish a historically-oriented, experience-based empirical version of economics in contrast to the prevailing Austrian “theoretical” economics. In the 1880s, the debate had focussed around Gustav Schmoller, the teacher of Max Weber and representative of the *younger historical school*, and his adversary Carl Menger. In his study *Investigating the Methods of the Social Sciences, especially Political Economy* (1883), Menger distinguished between theoretical, historical and practical sciences: he himself spoke for the theoretical realm which he further differentiated into an empirical-realistic and an “exact” position. The latter was supposed to work mathematically and to find social laws analogous to physical laws. Schmoller responded with his article *Methodology of the political and the social sciences* in which he accused Menger of dogmatism, of producing only hypothetical statements derived from other hypothetical statements, and of maintaining only an appearance of exactness and hard science.

Max Weber (1864-1920) rejected the proposition that rationalism is the only candidate for scientific, objective and valid knowledge claims and that all other approaches to knowledge were still proto-scientific and imperfect. As his contemporary Schumpeter put it, Weber saw “no objection of principle to what economic theorists actually did, though he disagreed with them on what they thought they were doing” (Schumpeter 1954: 819)¹. Instead Weber looked for an approach that was also able to fulfil other tasks: “The question as to how far, for example, contemporary “abstract theory” should be further elaborated is ultimately also a question of the strategy of science, which must, however, concern itself with other problems as well” (Weber 1949: 89).

Although he was Schmoller’s pupil, Weber was no proselyte of the historical school. While he agreed that “a science of man, and that is what economics is, inquires above all into the quality of man who are brought up in those economic and social conditions of existence” (Weber: 1980: 437), and that, in his eyes, “the science of economic policy is a political science” (Weber 1980: 438), he criticised the position of the historical school for its epistemological and methodological deficiencies (Weber 1975). In his eyes, the historical school lacked a clear terminology and an exactness in

¹ For Schumpeter, Max Weber was an “intellectual leader”, a “prince of science“, a “knight in shining moral armour”, and a polymath of unparalleled erudition. In Schumpeter’s view, his main achievement was to have transcended the limits of the German Historical School, inaugurating a “scientific” approach to history” (Osterhammel 1987: 107). Later, Schumpeter himself “was confronted with the alternatives “theory or history” that lingered on, even though the methodological dispute between Menger and Schmoller had cooled down by the turn of the century. From the start, Schumpeter refused to takes sides” (Osterhammel 1987: 108).

² In his study *Varieties of Social Economics: Joseph A. Schumpeter and Max Weber*, Osterhammel investigates the commonalities and differences between the two authors concerning this methodological dispute: “Both are opposed to Carl Menger’s ‘realist’ or ‘Aristotelian’ views, according to which the ‘laws’ of economics, analogous to those of the natural sciences, possess an objective existence.[...]. Weber and Schumpeter both favour a ‘nominalist’ position, but for different reasons. Weber is indebted to Neo-Kantian epistemology, whereas Schumpeter is influenced by Henry Poincaré’s ‘conventionalism’” (Osterhammel 1987: 110f).

concept formation, was too wedded to an unclear notion of “values” and could not show how eclectically heaping case study on case study would improve scientific understanding of the economy³.

Weber’s own solution was that there is a legitimate scientific interest in single cases, empirical detail and historical knowledge. Although he believed that there is an area of science where laws are of lesser interest, he was not satisfied with releasing this area from epistemological and methodological requirements. His solution is both historically interesting, because at the beginning of the 20th century Weber’s methodology was instrumental in securing sociology as an academic discipline, and epistemologically interesting, because of its relevance to the methodology of social simulation.

2.3 The Nomothetic and Ideographic Worlds in Neo-Kantianism

The position held by Kant and the Neo-Kantian school of philosophy in the rationalism/empiricism debate mentioned above offered a solution to the *Methodenstreit* for Weber. The disparity between the law-oriented, mathematically-based natural sciences and ‘history’ (meaning the historical sciences, the science of history, and historiography, the academic treatment of history as an object) was an important topic for Weber’s contemporaries in the philosophy of science. From them, Weber borrowed concepts to establish the academic and scientific claims of early sociology. Kant himself had distinguished between (natural) scientific knowledge and historical knowledge. According to Oakes (1987: 437), he argued that natural science is justified by the rationalist approach to “discover the necessary and universally valid laws that account for the properties of the phenomenal world. Historical claims to knowledge, on the other hand, are particular and contingent. In the final analysis, history has an inferior cognitive status, because it fails to qualify as a science according to the criteria for scientific knowledge that Kant elaborates in his *Critique of Pure Reason*: historical propositions lack the necessity and general validity that would qualify them as possible objects of scientific knowledge”.

The philosophers of the *Southwest German School of Neo-Kantianism* accepted Kant’s distinctions about legitimate knowledge claims, but tried to find a sound foundation for historical knowledge. Wilhelm Windelband (1848-1915) in *History and Natural Science* (*Geschichte und Naturwissenschaft*) and Heinrich Rickert (1863-1936) in *The Limits of Concept Formation in Natural Science* (*Die Grenzen der naturwissenschaftlichen Begriffsbildung*) established the notions “nomothetic” (law-making and law-finding) for the natural sciences and “ideographic” (individualistic and descriptive) for the historical sciences (cf. Windelband 1915: 144ff) following Kant’s original distinction. However, they did not accept that historical knowledge had an inferior cognitive status.

For Windelband, the “ideographic” feature of historical knowledge necessarily implied a special social science methodology. Oakes’ resumé of Windelband’s position is so exemplarily clear that we repeat it here at length:

³ Some proponents of the historical school were even more radical: “Everything concerned with exact notions and formal correctness is restricted to the natural sciences. Sciences of history, economy, law, literature and society, however, have to avoid this abuse because it leads them astray” (Prewo 1979: 16f, own translation).

“This theoretical interest [in the particular, contingent and individually significant] cannot be satisfied by natural science, which abstracts from the unique and qualitatively distinctive properties of real phenomena in order to disclose the laws on which they depend. This is the sense in which natural science is nomothetic. It has no intrinsic interest in the individual events of concrete reality. On the contrary, the individual datum is relevant to natural science only to the extent that it can be represented as a type, an instance of a generic concept, or a case that can be subsumed under a general law. This is a consequence of the ultimate theoretical purpose of natural science, which is to produce a system of maximally abstract and general laws, nomological regularities that govern all events. Nomothetic knowledge, therefore, represents the triumph of abstract thought over our perception of concrete reality. The interest of historical science, on the other hand, is idiographic. Here the purpose of knowledge is to comprehend the distinctive properties of the unique event itself. [...] The occurrence of individual events cannot be explained by general laws. Put another way, there is no set of nomological statements, regardless of how exhaustive and precise, from which any description of an individual event can be deduced. This is why our theoretical interest in individual phenomena cannot be satisfied by natural science. As Windelband claims, nomothetic and idiographic cognitive interests are independent and juxtaposed to one another: The law and the event remain as the ultimate and incommensurable entities of our world view. But if natural science cannot establish knowledge of individual phenomena, how is such knowledge possible? [...] Even though individual existence is subsumed under laws, it does not follow from them. Therefore, the law and reality, or concepts and reality, are incommensurable quantities” (Oakes 1987: 437ff)

2.4 Max Weber’s Foundations for the Social Sciences

This is the point where Weber took up the argument in order to derive his own position in this methodological dispute. He did not simply adopt either nomothetic or idiographic concepts. Windelband had already noted that the difference between law-based knowledge and historical knowledge did not emerge from a difference in the object (*nature* versus *culture*): one and the same object could be considered nomothetically and idiographically. The differences came from the epistemic aims and interests of the sciences – differences in aims and interests that remain today.

Some sciences are especially concerned with objects of cultural significance and meaning (e.g. modern Western capitalism, Greek democracy, Roman law, the French revolution, the Lisbon agenda etc.) and with people acting with reference to these meaningful objects and frameworks. It is certainly possible to deal with these objects nomothetically – not by using them in conjunction with universal quantors (e.g. “all Lisbon agendas are x ”), but as examples for more general statements (e.g. “all agendas are x , as illustrated by the Lisbon agenda”). Although it is easy to think of epistemological problems connected to such a statement, it is certainly meaningful and does carry some useful knowledge. However, the sciences concerned with objects of cultural significance and meaning normally have more ambitious epistemic interests.

In the sciences concerned with empirical *social* reality, the aim is to reconstruct human actions scientifically according to the significance and meaning assigned to them (a method called *verstehen*). This is the special advantage that the social sciences can offer. However, it has to rely on some idiographic features. The significant and meaningful objects we refer to (e.g. western capitalism) are related to culture and history. They are – as Weber called them – “historical individuals”, single cases. Investigating them in great detail gives us an understanding of the regularities we observe in social reality. It explains the *why* and *how* of social phenomena. As Weber puts it:

“Social science is an empirical science of experienced reality. The aim is to understand social reality that surrounds us in its peculiar character – on the one hand the contemporary framework and cultural meanings of all the single phenomena we observe now, and on the other hand, the reasons for their historical path that led to their special characteristics” (Weber 1988: 170f, own translation).

Cultural significance and meaning are not only important to explain social phenomena, but are also the general selection mechanism that all social knowledge production relies on for object formation. Following Kant’s argument that the *object per se* cannot be perceived, but just its *appearance for us*, objects for the social sciences are only available for study if they have cultural meaning and significance. Everything else is just noise. The mechanism of object formation guarantees that the social sciences are not concerned with a fixed combination of raw material but with a constant formation process of permanently changing objects that has the effect of making each of the objects individual.

A second individualistic feature of the social sciences relates to the science of action. An acting individual assigns meaning and significance to their actions according to their culture, history, socialisation experience, educational background and so on, and each “subjectively assigned meaning” forms a single case. Weber is interested in individual actions, not for their own sake, but to derive typologies and learn about regularities resulting from social and cultural frameworks such as institutions.

These idiographic features of Weber’s foundation for the social sciences did not exclude the nomothetic tradition. Weber also relied on the Neo-Kantian philosopher Heinrich Rickert, who had developed Windelband’s distinctions in a methodological direction. If cultural significance and meaning are the selection mechanism for permanently changing social science object formation, then the notions and concepts to name these individual objects and their relations cannot claim to represent social reality *per se*, as they would have to if they were to be a pre-requisite for law statements. Instead, concept formation has to match object formation. The objects of social science are not representations of some natural kinds, they are *constructed* by applying a selection mechanism. The concepts, accordingly, are not images of reality that map representations and they have to follow the same fluent construction mechanism as applied in object formation. Concept formation proceeds “individualistically”.

This is achieved by the Weberian ideal type. An ideal type is an instrument of concept formation “that is specific and indispensable for any sciences concerned with human culture” (Weber 1988: 190). To form an ideal type, a conceptual extreme is constructed from empirical reality. This gives it a logical consistency that enables it to serve as a formal instrument and heuristic tool. “The ideal type is constructed by

partially emphasising one or more characteristics and by combining many diffuse and discrete, more or less - sometimes not at all – present individual phenomena – a procedure which allows one to build a unified thought concept from these partially emphasised characteristics. In its conceptual purity the ideal type does not exist in empirical reality, it is a utopia. For historical work the task remains to state for each single case how close or how distant reality comes to this thought concept” (Weber 1988: 191). An ideal type is a kind of heuristic jig pattern: with this descriptive tool we can describe empirical reality, develop typologies, and characterise single objects, regularities and processes. The value of an ideal type depends on its utility: if an ideal type does not assist research it should be modified or even abandoned⁴.

This method of concept formation permits a very exact account of the way that scientific statements are generated. The ‘recipe’ of concept formation is completely transparent. One no longer needs to consider ontological or epistemological problems such as references to reality or classification hierarchies. Instead, one can remain on the level of scientific logic: whosoever follows the concept formation procedure will arrive at the same descriptions, conclusions and results. This is the nomothetic implication in Weber’s solution. Exact notions and precise formal criteria for scientific statements are as important in the social sciences as in the natural sciences. This is the way that the social sciences concerned with single cases using single concepts can produce results that are universally valid for everybody, everywhere (Weber 1988: 156).

2.5 Epistemic Aims and Interests – A Continuum

Weber (e.g. 1988: 12ff) describes what mathematics can offer the social sciences and where law-finding strategies can be useful, but he also indicates where and why a different approach is a necessary complement in order to provide additional knowledge⁵. The social sciences are not a ‘pre-science’ waiting to approximate the state of the natural sciences via more and more discovery and mathematisation of the laws of the social realm. Weber’s solution should have finalised the methodological dispute in political economy: as Joseph Schumpeter put it, it was “substantially a history of wasted energies which could have been put to better use” (1954: 814). But, since a similar debate in social simulation between the proponents of “abstract social processes” and those of “history-friendly and stakeholder-sensitive models” continues, it might be useful to see what lessons we can learn from the dispute.

If we leave aside the sometimes colonising and patronising rhetoric of “good science” in the present debate about social simulation, the Weberian view makes plain the differences in the epistemic aims and interests that lead to different – but nevertheless scientifically sound – simulation methodologies.

⁴ With this approach, Weber again was in accordance with Joseph Schumpeter: “To claim that the ‘laws’ of economic theory were nothing but “hypotheses made up by us”, “just arbitrary as definitions” and only to be judged in terms of their “utility” was to fly in the face of Vienna orthodoxy. [...] Weber was a methodological constructivist, and Schumpeter, as one of the very few economists, joined him in this at an early stage” (Osterhammel 1987: 111).

⁵ Weber agreed that there were law-like regularities in the social world that could be expressed mathematically but also argued that the abstract level of such statements missed the object and the scientific task of the social sciences.

The nomothetically dominated sector is characterised by an interest in the most abstract and general features of the social world and the mathematically tractable axioms and theorems that surround them (left side of figure 1). A law in this sector has to be well-formed (a scientific hypothesis) and must provide generality:

“that is, we demand that at least one of the variables occurring in the law formula be prefixed by the operator ‘for every’, ‘for almost every’, or ‘for most’; if the former is the case, i.e. if the law is a strict universal hypothesis, then we usually dispense with the explicit addition of the quantifier. If the law refers to an individual (as in the case of the geophysical laws, which refer to our planet), we require that the statement expresses the regular behaviour of the individual as shown by a universal quantifier with respect to time; the quantifier may be unrestricted or bounded, and it may be manifest or tacit but must be there: otherwise the proposition would be singular rather than general. If, on the other hand, the law formula does not refer to an individual but to a class, we may tolerate quasi-generality (...). ‘Most’ and ‘almost all’ are not treated with respect by logicians, who lump them together with ‘there is at least one’, but in science their status is much higher than that of the existential operator: an almost-all-formula may be a law proper, and a most-formula may promise a universal law” (Bunge 1967: 334f).

Single cases are not excluded in this perspective, as Bunge notes. However, it is important in what way they are of interest. Either they are just examples: the aim is to arrive at law statements for classes of objects, the closer to universal the better, and single cases just instantiate the law. Or, if single cases are treated as individuals, they must provide the required generality in the time dimension.

Bunge argues that, though this approach relies on a realist ontology, it only needs epistemological consideration: statements about laws are something other than the underlying structures that the statements refer to. Science only refers to statements:

“‘Law’ (or ‘objective law’, or ‘nomic structure’) designates an objective pattern of a class of facts (things, events, processes), i.e. a certain constant relation or mesh of constant relations really obtaining in nature, whether we know it or not. A law, in this sense of nomic structure, is an extraconceptual object, like the flow of a river. But, unlike the flow of a river, its laws cannot be pointed too: they are imperceptible. In short: the concept of objective law is empirically meaningless: which shows that it is not trivial. We cannot exhibit a specimen of an objective law but we can utter a definite description (...). Whether we acknowledge the existence of objective laws or not we need this concept, if only to argue against the philosophical hypothesis that there are laws underlying the law statements.

‘Law formula’ (or ‘nomological statement’) designates a proposition or a propositional function that is usually supposed to describe a law or a part of a law (nomic structure). A law formula is a conceptual object, namely a scientific hypothesis satisfying certain requisites of generality, corroboration, and systematicity” (Bunge 1967: 344).

The idiographically dominated area is characterised by an interest in the particular and singular features of social phenomena and the detailed descriptions and history that contextualise them (right side of figure 1). Single cases are of special interest for their own sake – not only as representative of features that are constant or generalisable over

time. On the contrary, the ever changing context, i.e. the exact time and location, is the most important determinant for each observable. The researcher acts like an anthropologist and ethnographer, reconstructing a social field in a permanently changing cultural context, where:

“we do not have direct access, but only that small part of it which our informants can lead us into understanding. This is not as fatal as it sounds, for, in fact, not all Cretans are liars, and it is not necessary to know everything in order to understand something. But it does make the view of anthropological analysis as the conceptual manipulation of discovered facts, a logical reconstruction of a mere reality, seem rather lame. To set forth symmetrical crystals of significance, purified of a material complexity in which they were located, and then attribute their existence to autogenous principles of order, universal properties of the human mind, or vast, *a priori* weltanschauungen, is to pretend a science that does not exist and imagine a reality that cannot be found” (Geertz 1973: 20).

Since object formation is part of the procedure, a constructivist approach to knowledge production is implied. However, again we need not necessarily refer to ontology: only epistemological issues such as theory formation using ideal types as heuristic tools are involved at the level of scientific discourse.

It is hard to envisage any empirical academic discipline implementing either extreme of the continuum. However, looking at the academic spectrum as a whole, *theoretical mathematics* and *narrative hermeneutical history* are located near the poles, and all the empirical sciences may be placed somewhere in between.

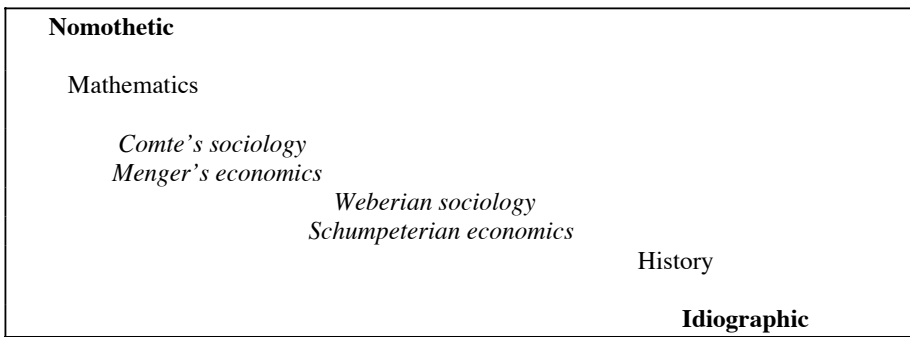


Fig. 1. Continuum of epistemological approaches, and the positions taken by some early social scientists

3 Nomothetic and Ideographic in Social Simulation

Using agent-based models, it is claimed that we can find “the trade-off between simplicity and abstracting in modelling, and [take] into account the complexity of (...) reality” (Pyka and Grebel 2006: 17). How is this trade-off realised according to the epistemological continuum visualised in Figure 1?

In some simulation studies, the term *social mechanism* is used as a synonym for *social law*, implying a nomothetic approach. However, as Mayntz states “a still incomplete list of definitions assembled by Mahoney counts 24 different definitions by 21 authors. Mechanisms are considered to be lawful processes, yet they are also opposed to laws” (Mayntz 2003: 3). One can find definitions which stress that mechanisms are “truncated abstract descriptions” generalising causal propositions (e.g. Machamer, Darden and Craver 2000: 15). However, one can also find the opposite where a social mechanism (SM) is a “well-articulated set of causes responsible for a given social phenomenon. With the exception of typical simple ones, SMs tend to be idiosyncratic and singular” (Boudon 1998: 172).

The term *mechanism* is most useful, not to distinguish between nomothetically and idiographically oriented simulation studies, but to point to commonalities. For Mayntz, a mechanism – as opposed to propositions about correlations⁶ - is a causal reconstruction of a social event, structure or development. “There should not be too much proximity between cause and effect”, “mechanisms state *how*, by what intermediate steps, a certain outcome follows from a set of initial conditions”, a process which “can contain feedback loops, and each unit involved in the process may undergo changes” (Mayntz 2003: 4f).

This idea of mechanism lies at the heart of social simulation. “Agent-based models provide computational demonstrations that a given microspecification is in fact sufficient to generate a macrostructure of interest. (...) Thus, the motto of generative social science, if you will, is: If you didn’t grow it, you didn’t explain its emergence“ (Epstein 2006: 8). All agent-based models involve more or less elaborated, complex, and detailed social mechanisms in this sense. The reason that modellers employ different degrees of generality is related to their epistemic aims and interests and can therefore be visualised on the continuum shown in Figure 1. Figure 2 gives a rough illustration of how one might use such a visualisation.

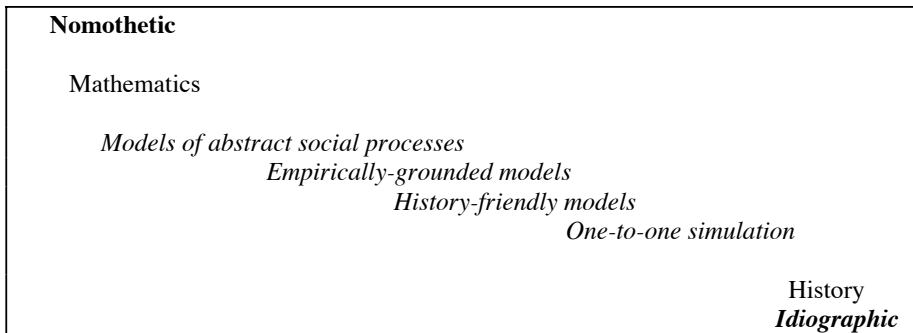


Fig. 2. A Typology of Simulation Studies

⁶ The same opposition applies to the covering-law model of causal explanation, by the way, as Mayntz states citing Bunge and Hedström & Swedberg: “The covering-law model is often criticized for the same reason brought against correlational analysis: a nomological-deductive explanation involving lawlike prepositions supplies no understanding, it gives no clue whatsoever as to why a relationship exists: covering-law explanations in the social sciences therefore normally are ‘black-box’ explanations” (Mayntz 2003: 3).

As an example, consider the general human concept of neighbourhood. There are nomothetically-oriented simulation studies that assume that there is a law that states that every neighbourhood interaction involves a transfer which depends on the attributes of the actors involved and the kind of contact. This is an idea taken from classical epidemiology in biology and applied to the sociology of neighbourhoods⁷. If we treat *epidemic* as an underlying abstract social process we can find simulation studies using this general “social law” (*you get what your neighbours have*) to model social phenomena such as innovation diffusion, segregation, and opinion dynamics (e.g. Bagni *et al.*, 2002; Deffuant, 2006; Dunham, 2005; Hegselmann & Krause, 2002; Huang *et al.*, 2005; Otter *et al.*, 2001; Salzarulo, 2006; Srbljinovic *et al.*, 2003; Stauffer *et al.*, 2004).

However, the criteria of nomothetic modelling are not completely met by these approaches. Neighbourhood dynamics modelling could be considered to be like a law statement “if and only if (i) it is general in some respect and to some extent; (ii) it has been empirically confirmed in some domain in a satisfactory way, and (iii) it belongs to some scientific system” (Bunge 1967: 361). Concerning the second requirement, often only anecdotal evidence is presented when targeting the empirical confirmation of epidemic social processes. The third requirement, however, is nearly always dispensed with. To belong to a scientific system, it would have to be embedded in a theory, i.e. refer to a given subject matter, in which every member of its set of participating scientific hypotheses is either an initial assumption (axiom, subsidiary assumption, or data) or a logical consequence of one or more initial assumptions, and which provides explanation and predictions that are testable (cf. Bunge 1967: 381ff).

We would have to consider carefully the epistemic claims of modelling “epidemic social processes” to see where and how these requirements are covered. So far, simulation studies that model neighbourhood interaction are usually rather unconnected to the body of existing social theory and unconcerned about empirical validation issues.

Looking at the right side of the continuum, we find completely different concerns. For idiographically oriented modellers, the aim is to *understand and explain a special case*, the history of an individual formation. For example, the modellers of the Kayenta Anasazi population in Long House Valley state: “Our model closely reproduces important spatial and demographic features of the Anasazi in Long House Valley from about A.D. 800 to 1300. To ‘explain’ an observed spatiotemporal history is to specify agents that generate - or grow - this history. By this criterion, our strictly environmental account of the evolution of this society during this period goes a long way toward *explaining* this history“ (Axtell *et al.* 2002: 7279).

Such idiographic modelling, which starts from data not from theory, inherits the problems of empiricism: it assumes the possibility of “innocent data” provided by pure observations that will in the end give the complete picture. “The base of empirical science becomes [...] a system of *a priori* theories. If their claims are false, it is only because they are inconsistent: if they are true, they are *a priori* true. Bed-rock theory-elements of empirical science would have exactly the same epistemological

⁷Maassen and Weingart (2000) identify the facilitation of cross-domain communications as one of the main advantages of transferring metaphors in science. However, to what extent such a conceptual transfer can successfully be completed is critically discussed (e.g. in the chapters of Weingart and Stehr 2000).

status as formal systems of logic.” (Balzer, Moulines and Sneed 1987: 417). We will not elaborate on the complex interactions of empiricism, constructivism, and relativism (cf. Ahrweiler 2001), but they certainly lurk behind the farthest end of the continuum’s idiographic end.

4 Evaluating Simulation Studies

As we have argued elsewhere (Ahrweiler and Gilbert 2005), a simulation is good when it serves our epistemic aims and interests. In this chapter, we have pointed out that there is not just one legitimate aim for simulation studies, but a continuum between the nomothetic and idiographic extremes. Whether particular studies are “good” or “bad” depends on the criteria defined by their particular user group, that is, the group defining their specific epistemic aims and interests. The evaluation of any simulation is guided by the expectations, anticipations and experience of the communities that use it.

Accordingly, there is a range of evaluation criteria, corresponding to the continuum of objectives. Although these criteria should rightfully be defined by the communities themselves, we can review attributes that we know to be of some importance for them.

For nomothetically oriented modellers, the *simplicity* of their models seems to be crucial. So, for example, after exulting that “Occam’s razor slashed again” (Kennedy and Eberhart 1996: section 3.7), the conclusion of a particle swarm simulation of social behaviour states:

“The goals in developing it have been to keep it simple and robust, and we seem to have succeeded at that. The algorithm is written in a very few lines of code, and requires only specification of the problem and a few parameters in order to solve it. This algorithm belongs ideologically to that philosophical school that allows wisdom to emerge rather than trying to impose it, that emulates nature rather than trying to control it, and that seeks to make things simpler rather than more complex. Once again nature has provided us with a technique for processing information that is at once elegant and versatile” (Kennedy and Eberhart 1996: section 6).

Where the epistemic interest points to an empirically-rich idiographic modelling strategy, criteria such as the model’s fits to data, its predictive adequacy, its practical usability and its transparency are important issues for evaluation. Sometimes, the hoped-for achievements can be argued quite strongly: the article *Towards Good Social Science* (Moss and Edmonds 2005) is introduced as “as propounding a neo-positivist position — that the social sciences should be more like the natural sciences”. In particular the authors argue for the following:

- “The fundamental *priority* of observation and evidence over models and theory
- The importance of the multiple validation of theory
- Not relying on theory that is not sufficiently validated (and certainly not as a justification for further theory generation)
- That evidence should not be excluded if at all possible, especially anecdotal evidence from stakeholders

- That much more effort should be expended towards developing new techniques for the observation of social phenomena
- That much descriptive modelling at a low, concrete level will probably be necessary before successful and useful more general theory can be developed
- That some of the modelling and description needs to be of a formal (but probably computational and not analytic) nature so that we know what we are comparing and talking about
- That agent-based simulation can help facilitate the above” (Moss and Edmonds 2005: 7.1-7.2)

In contrast, we have suggested that this is only one of a range of possible and justifiable positions.

Summarising, the debate within social simulation has many connections with the debates that have echoed down the years within the wider social science community about the character of social science knowledge and the appropriate methodological assumptions on which social science research should rest. Learning from this debate and its resolution, evaluation criteria for simulation studies need to be tuned to their research aims, methodological positions and domains of study. Communities have to offer and develop explicitly their own “best practices” of doing simulation research, rather than having their epistemological approach dictated to them by methodologists.

References

1. Ahrweiler, P., Gilbert, N.: *Caffè Nero: the Evaluation of Social Simulation*. *Journal of Artificial Societies and Social Simulation* 8(4), 14 (2005), <http://jasss.soc.surrey.ac.uk/8/4/14.html>
2. Axelrod, R.: *Advancing the Art of Simulation in the Social Sciences*. In: Conte, R., Hegselmann, R., Terna, P. (eds.) *Simulating Social Phenomena*, pp. 21–40. Springer, Heidelberg (1997)
3. Axtell, R.L., Epstein, J.M., Dean, J.S., Gumerman, G.J., Swedlund, A.C., Harburger, J., Chakravarty, S., Hammond, R., Parker, J., Parker, M.: *Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley*. In: *PNAS*, May 14, 2002, vol. 99 (suppl. 3), pp. 7275–7279 (2002), http://www.pnas.org/cgi/content/full/99/suppl_3/7275?ck=nck
4. Bagni, R., Berchi, R., Cariello, P.: *A Comparison of Simulation Models Applied to Epidemics*. *Journal of Artificial Societies and Social Simulation* 5(3), 5 (2002), <http://jasss.soc.surrey.ac.uk/5/3/5.html>
5. Beck, N., Katz, J.: *What to Do (and not to Do) with Time-Series Cross-Section Data*. *American Political Science Review* 89, 634–647 (1995)
6. Boudon, R.: *La logique du social*. Hachette, Paris (1979)
7. Boudon, R.: *La place du desordre*. Presses Universitaires de France, Paris (1984)
8. Boudon, R.: *Social Mechanisms Without Black Boxes*. In: Hedström, P., Swedberg, R. (eds.) *Social Mechanisms. An Analytical Approach to Social Theory*, pp. 172–203. Cambridge University Press, Cambridge (1998)
9. Bunge, M.: *Scientific Research I. The Search for System*. Springer, New York (1967)
10. Bunge, M.: *Mechanisms and Explanation*. *Philosophy of the Social Sciences* 27(4), 410–465 (1997)

11. Büthe, T.: Taking Temporality Seriously: Modelling History and the Use of Narratives as Evidence. *American Political Science Review* 96(3), 481–493 (2002)
12. Calhoun, C.: Explanation in Historical Sociology: Narrative, General Theory, and Historically Specific Theory. *American Journal of Sociology* 104(3), 846–871 (1998)
13. Coleman, J.S.: Introduction to Mathematical Sociology. Free Press, New York (1964)
14. Coleman, J.S.: Social Theory, Social Research, and a Theory of Action. *American Journal of Sociology* 91(6), 1309–1335 (1986)
15. Coleman, J.S.: Foundations of Social Theory. Harvard University Press, Cambridge (1990)
16. Comte, A.: Die Soziologie. Positive Philosophie. Stuttgart, Kroener (1975)
17. Craver, C.: Role Functions, Mechanisms, and Hierarchy. *Philosophy of Science* 68(1), 53–74 (2001)
18. David, P.: Clio and the Economics of QWERTY. *American Economic Review* 75, 332–337 (1985)
19. Deffuant, G.: Comparing Extremism Propagation Patterns in Continuous Opinion Models. *Journal of Artificial Societies and Social Simulation* 9(3), 8 (2006), <http://jasss.soc.surrey.ac.uk/9/3/8.html>
20. Dilthey, W.: Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften. Frankfurt (The structure of the historical world in the humanities) (1970)
21. Dunham, J.B.: An Agent-Based Spatially Explicit Epidemiological Model in MASON. *Journal of Artificial Societies and Social Simulation* 9(1), 3 (2005), <http://jasss.soc.surrey.ac.uk/9/1/3.html>
22. Elias, N.: Über den Prozeß der Zivilisation, 2nd edn. Francke AG, Bern (1969)
23. Elster, J.: Nuts and Bolts for the Social Sciences. Cambridge University Press, Cambridge (1989)
24. Elster, J.: A Plea for Mechanisms. In: Hedström, P., Swedberg, R. (eds.) *Social Mechanisms. An Analytical Approach to Social Theory*, pp. 45–73. Cambridge University Press, Cambridge (1998)
25. Epstein, J.M.: Generative Social Science. Studies in Agent-Based Computational Modeling. Princeton University Press, Princeton (2007)
26. Esser, H.: Soziologie. Allgemeine Grundlagen. Frankfurt, Campus (1993)
27. Esser, H.: Was könnte man (heute) unter einer Theorie mittlerer Reichweite verstehen? In: Mayntz, R. (ed.) *Akteure - Mechanismen - Modelle. Zur Theoriefähigkeit makro-sozialer Analysen*, pp. 128–150. Frankfurt, Campus (2002)
28. Franks, D.W., Kaufmann, P., Stagl, S.: Simulating the Adoption of Organic Farming Practices in New EU Member States – A Study of Latvia. SPRU 40th Anniversary Conference Paper (2006), <http://www.sussex.ac.uk/Units/spru/events/ocs/viewpaper.php?id=294>
29. Geertz, C.: Thick Description: Toward an Interpretive Theory of Culture. In: *The Interpretation of Cultures: Selected Essays*, pp. 3–30. Basic Books, New York (1973)
30. Glennan, S.S.: Mechanisms and the Nature of Causation. *Erkenntnis* 44, 49–71 (1996)
31. Granovetter, M.: Threshold Models of Collective Behavior. *American Journal of Sociology* 83(6), 1420–1443 (1978)
32. Habermas, J.: Ein anderer Ausweg aus der Subjektphilosophie: Kommunikative vs. Subjektzentrierte Vernunft. In: Habermas, J. (ed.) *Der philosophische Diskurs der Moderne. Vorlesungen*, vol. 12, pp. 344–379. Frankfurt/M.: Suhrkamp, S. (1985)
33. Hall, P.A.: Aligning Ontology and Methodology in Comparative Research. In: Mahoney, J., Rueschemeyer, D. (eds.) *Comparative Historical Research in the Social Sciences*. Cambridge University Press, New York (in print) (2003)

34. Hedström, P., Swedberg, R. (eds.): *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge University Press, Cambridge (1998)
35. Hedström, P., Swedberg, R.: *Social Mechanisms*. *Acta Sociologica* 39(3), 281–308 (1996)
36. Hegselmann, R., Krause, U.: *Opinion Dynamics and Bounded Confidence Models. Analysis and Simulation*. *Journal of Artificial Societies and Social Simulation* 5(3), 2 (2002), <http://jasss.soc.surrey.ac.uk/5/3/2.html>
37. Hempel, C.G.: *Aspects of Scientific Explanation - And Other Essays in the Philosophy of Science*. Free Press, New York (1965)
38. Hoover, K.D.: *The Logic of Causal Inference. Econometrics and the Conditional Analysis of Causation*. *Economics and Philosophy* 6, 207–234 (1990)
39. Huang, C.-Y., Sun, C.-T., Lin, H.-C.: *Influence of Local Information on Social Simulations in Small-World Network Models*. *Journal of Artificial Societies and Social Simulation* 8(4), 8 (2005), <http://jasss.soc.surrey.ac.uk/8/4/8.html>
40. Johnson, J.: *How Conceptual Problems Migrate: Rational Choice, Interpretation, and the Hazards of Pluralism*. *Annual Review of Political Science* 5, 223–248 (2002)
41. Kaesler, D.: *Max Weber. An Introduction to his Life and Work*. Polity Press, Cambridge, University of Chicago Press, Chicago (1988)
42. Karlsson, G.: *Social Mechanisms. Studies in sociological theory*. Almqvist & Wiksell, Stockholm (1958)
43. Kennedy, J., Eberhart, R.: *Particle Swarm Optimization*. In: *Proc. IEEE Int'l. Conf. on Neural Networks (Perth, Australia)*, IEEE Service Center, Piscataway, NJ, pp. IV:1942–IV:1948 (1995), <http://www.engr.iupui.edu/~shi/Coference/psopap4.html>
44. Kertész, A.: *Artificial intelligence and the sociology of scientific knowledge. Prolegomena to an integrated philosophy of science*. Frankfurt/M, Lang. (1993)
45. Kiser, E., Hechter, M.: *The Debate on Historical Sociology: Rational Choice and Its Critics*. *American Journal of Sociology* 104(3), 785–816 (1998)
46. Kitschelt, H.: *Accounting for Postcommunist Regime Diversity: What Counts as a Good Cause?* In: Ekiert, G., Hanson, S. (eds.) *Capitalism and Democracy in Eastern and Central Europe. Assessing the Legacy of Communist Rule*. Cambridge University Press, Cambridge (in print) (2003)
47. Knight, J.: *Models, Interpretations, and Theories: Constructing Explanations of Institutional Emergence and Change*. In: Knight, J., Sened, I. (eds.) *Explaining Social Institutions*. University of Michigan Press, Ann Arbor (1995)
48. Kuran, T.: *Sparks and Prairie fires: A theory of unanticipated political revolution*. *Public Choice* 61, 41–74 (1989)
49. Little, D.: *Varieties of Social Explanation. An Introduction to the Philosophy of Social Science*. Westview Press, Boulder (Col.) (1991)
50. Little, D.: *On the Scope and Limits of Generalizations in the Social Sciences*. *Synthese* 97, 183–207 (1993)
51. Maasen, S., Weingart, P.: *Metaphors and the Dynamics of Knowledge*. Routledge, London (2000)
52. Machamer, P., Darden, L., Craver, C.F.: *Thinking about Mechanisms*. *Philosophy of Science* 67(1), 1–25 (2000)
53. Mahoney, J.: *Beyond Correlational Analysis: Recent Innovations in Theory and Method*. *Sociological Forum* 16(3), 575–593 (2001)
54. Mayntz, R.: *Mechanisms in the Analysis of Macro-Social Phenomena*. MPI Working Paper 03/3 (April 2003); re-printed in *Philosophy of the Social Sciences* 34(2), pp. 237–259 (2004)

55. Mayntz, R.: Soziale Diskontinuitäten. Erscheinungsformen und Ursachen. In: Soziale Dynamik und politische Steuerung. Theoretische und methodologische Überlegungen, pp. 115–140. Frankfurt/Main, Campus (1988/1997)
56. Mayntz, R.: Zur Theoriefähigkeit makro-sozialer Analysen. In: Mayntz, R. (ed.) Akteure - Mechanismen - Modelle. Zur Theoriefähigkeit makro-sozialer Analysen, pp. 7–43. Frankfurt/Main, Campus (2002)
57. McAdam, D., Tarrow, S., Tilly, C.: Dynamics of Contention. Cambridge University Press, Cambridge (2001)
58. Menger, C.: Untersuchungen über die Methode der Sozialwissenschaften und der politischen Ökonomie insbesondere. Leipzig (Investigating the methods of the social sciences, especially political economy) (1883)
59. Merton, R.K.: On Sociological Theories of the Middle Range. In: On Sociological Theory. Five Essays, Old and New, pp. 39–72. Free Press, New York (1957)
60. Moss, S., Edmonds, B.: Towards Good Social Science. Journal of Artificial Societies and Social Simulation 8(4), 13 (2005), <http://jasss.soc.surrey.ac.uk/8/4/13.html>
61. Müller, H.-P.: Soziologie in der Eremitage? Skizze einer Standortbestimmung. In: Barlösius, E., Müller, H.-P., Sigmund, S. (eds.) Gesellschaftsbilder im Umbruch. Leske + Budrich, Opladen (2001)
62. Nagel, E.: The Structure of Science. Problems in the Logic of Scientific Explanation. Routledge & Kegan Paul, London (1961)
63. Nedelmann, B., Mayntz, R.: Eigendynamische soziale Prozesse. Anmerkungen zu einem analytischen Paradigma. Kölner Zeitschrift für Soziologie und Sozialpsychologie 39, 648–668 (1987)
64. Oakes, G.: Weber and the Southwest German School: The Genesis of the Concept of the Historical Individual. In: Mommsen, W., Osterhammel, J. (eds.) Max Weber and his Contemporaries, pp. 434–446. Allen & Unwin, London (1987)
65. Osterhammel, J.: Varieties of Social Economics: Joseph A. Schumpeter and Max Weber. In: Mommsen, W., Osterhammel, J. (eds.) Max Weber and his Contemporaries, pp. 106–120. Allen & Unwin, London (1987)
66. Ostrom, E.: Governing the Commons. The Evolution of Institutions for Collective Action. Cambridge University Press, Cambridge (1990)
67. Otter, H.S., van der Veen, A., de Vriend, H.J.: ABLOoM: Location Behaviour, Spatial Patterns, and Agent-Based Modelling. Journal of Artificial Societies and Social Simulation 4(4), 2 (2001), <http://jasss.soc.surrey.ac.uk/4/4/2.html>
68. Pawson, R.: Middle-range realism. Archives Européennes de Sociologie XLI (2), 283–324 (2000)
69. Petersen, R.: Mechanisms and structures in comparison. In: Bowen, J. (ed.) Critical Comparisons in Politics and Culture. Cambridge University Press, Cambridge (1999)
70. Pierson, P.: Increasing Returns Path Dependence, and the Study of Politics. American Political Science Review 94, 251–267 (2000)
71. Prewo, R.: Max Webers Wissenschaftsprogramm. Suhrkamp (Max Weber's Programme of Science), Frankfurt (1979)
72. Pyka, A., Grebel, T.: Agent-based Modelling – A Methodology for the Analysis of qualitative Development Processes. In: Billari, F.C., Fent, T., Prskawetz, A., Scheffran, J. (eds.) Agent-Based Computational Modelling. Applications in Demography, Social, Economic and Environmental Sciences. Physica, Heidelberg (2006)
73. Ragin, C.C.: The Comparative Method. Moving Beyond Qualitative and Quantitative Strategies. University of California Press, Berkeley (1987)

74. Ragin, C.C.: *Fuzzy-Set Social Science*. Chicago University Press, Chicago (2000)
75. Rickert, H.: *Die Grenzen der naturwissenschaftlichen Begriffsbildung*. Tuebingen (The limits of concept formation in the natural sciences) (1902)
76. Salzarulo, L.: A Continuous Opinion Dynamics Model Based on the Principle of Meta-Contrast. *Journal of Artificial Societies and Social Simulation* 9(1), 13 (2006), <http://jasss.soc.surrey.ac.uk/9/1/13.html>
77. Schmoller, G.: Zur Methodologie der Staats- und Sozialwissenschaften. In: *Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reich* 7, pp. 239–258 (1883) (Methodology of the political and the social sciences)
78. Schnaedelbach, H.: *Rationalitaet*. Philosophische Beitrage. Suhrkamp, Frankfurt (1984) (Rationality. Philosophical Contributions)
79. Schumpeter, A.: *History of Economic Analysis*, London (1954)
80. Somers, M.R.: We're no Angels: Realism, Rational Choice, and Relationality in Social Science. *American Journal of Sociology* 104(3), 722–784 (1998)
81. Srblijinovic, A., Penzar, D., Rodik, P., Kardov, K.: An Agent-Based Model of Ethnic Mobilisation. *Journal of Artificial Societies and Social Simulation* 6(1), 1 (2003), <http://jasss.soc.surrey.ac.uk/6/1/1.html>
82. Stauffer, D., Sousa, A., Schulze, C.: Discretized Opinion Dynamics of the Deffault Model on Scale-Free Networks. *Journal of Artificial Societies and Social Simulation* 7(3), 7 (2004), <http://jasss.soc.surrey.ac.uk/7/3/7.html>
83. Stinchcombe, A.L.: The Conditions of Fruitfulness of Theorizing about Mechanisms in Social Science. *Philosophy of the Social Sciences* 21, 367–388 (1991)
84. Stinchcombe, A.L.: Monopolistic competition as a mechanism: Corporations, universities, and nation-states in competitive fields. In: Hedström, P., Swedberg, R. (eds.) *Social Mechanisms. An Analytical Approach to Social Theory*, pp. 267–305. Cambridge University Press, Cambridge (1998)
85. Thagard, P.: *Computational Philosophy of Science*. MIT Press, Cambridge (1988)
86. Thelen, K.: Historical Institutionalism in Comparative Politics. *The Annual Review of Political Science* 2, 369–404 (1999)
87. Tilly, C.: Mechanisms in Political Processes. *Annual Review of Political Science* 4, 21–41 (2001)
88. von Wiese, L.: *System der allgemeinen Soziologie: Beziehungslehre - als Lehre von den sozialen Prozessen und den sozialen Gebilden des Menschen*. Duncker und Humblot, Berlin (1933)
89. Weber, M.: *The Methodology of the Social Sciences* (translated and edited by E.A. Shils and H.A. Finch), New York (1949)
90. Weber, M.: *The National State and Economic Policy* (translated by B. Fowkes). *Economy and Society* 9, 428–449 (1980)
91. Weber, M.: *Roscher and Knies: The logical Problems of Historical Economics* (translated by G. Oakes), New York (1975)
92. Weber, M.: *Gesammelte Aufsätze zur Wissenschaftslehre*. Tuebingen: UTB/Mohr. (Collected articles concerning the methodology of the social sciences) (1988)
93. Weingart, P., Stehr, N. (eds.): *Practising Interdisciplinarity*. University of Toronto Press, Toronto (2000)
94. Windelband, W.: *Geschichte und Naturwissenschaft*. In: Windelband, W. (ed.) *Praeludien* 2, Tuebingen, pp. 136–160 (1915) (History and Natural Science)

From Simulation to Theory (and Backward)

Rosaria Conte

LABSS - Laboratory of Agent-Based Social Simulation
ISTC/CNR Rome, Italy
rosaria.conte@istc.cnr.it

Abstract. Of late, due to the perceived advantages of the generative paradigm (Epstein 2002, 2005; Arthur 2004), a generative variant of agent based modeling and simulation, i.e. Agent Based Generative Social Simulation (ABGSS), has received a great impulse. In this paper, weak ABGSS, i.e. the thesis that growing a social effect is necessary but insufficient to explain it, is supported. Casting a critical eye on the debate about generative simulation, it will be argued that ABGSS needs to be fed by, but at the same time provides feedback to, two theoretical complements which must be formulated prior and independent of simulation itself: (a) bottom-up theory of general local rules, and of the process from them to macroscopic effects; (b) theory of downward causation, showing how local rules are modified by the effects they contribute to achieve. This twofold thesis will be carried out while discussing three main examples of social phenomena: the witness effect, Schelling's segregation model and the ethnic homogeneity of violence, and the minority game.

Keywords: Agent-based generative social simulation, bottom-up theory, top-down theory, multiple realizability, social emergence.

1 Introduction

In the last few years, Agent Based Social Simulation (ABSS) has received a great impulse, owing among other reasons to the perceived advantages of the generative paradigm (Epstein, 2002; 2005; Arthur, 2004).

Epstein formulated a variant of agent based social simulation, i.e. Agent Based *Generative* Social Simulation (ABGSS), which he explicitly asserts to be necessary for explanation:

For all x (not Gx materially implies not Ex)

In words, if you do not grow something, you cannot explain it. However, the author warns the reader that since there may be several paths leading to the same result, growing x is said to be not sufficient to explain it¹. Hence, two theses about ABGSS emerge:

- *Weak:* Grow is necessary but not sufficient to Explain ($Ex \supset Gx$),
- *Strong:* Grow is not only necessary, but also sufficient to Explain; Grow = Explain ($Gx = Ex$)

¹One ought to be clear in keeping distinct the approach, which is said to be necessary but insufficient, from the explanation it allows, precisely because the latter is said to be sufficient but not necessary.

Epstein supports the weak thesis.

In this paper, I will argue in favour of weak ABGSS for reasons which overlap only in part with Epstein’s. Casting a critical eye on the debate about generative simulation, I will argue that ABGSS is an innovative process crucial to produce effects upon which to theorize, in principle allowing to open the black box during behavioural experimentation. However, is no theory *per se*: not (only) because, as Epstein recognizes, there may be several ways to obtain the *explanandum*, but also because:

Often, generation turns into a *shortcut*: much as happened for a long time with experimental behavioural sciences, the general precept “what are the minimal conditions under which x occurs” leads to *ad hoc*² analytical as well as simulation models. Hence, a bottom-up generative theory is needed to generate the *explanandum* in a scientifically significant way.

In the variant formulated by Epstein, ABGSS is only a *bottom-up* paradigm, ignoring downward causation and the micro-macro link. This limits the explanatory and, for that matter, generative power of simulation. A generative theory of downward causation is also needed, showing how local rules are modified by the effects they contribute to achieve.

In other words, generative simulation is an invaluable tool for explanation if it is both *theory-making* (from simulation to theory) and *theory-driven* (from theory back to simulation). I will show this by some examples: the witness effect, Schelling’s segregation model and the ethnic homogeneity of violence, and the minority game.

1.1 Organization of the Paper

In the first section, both notions of generation and explanation will be re-examined. In the following two sections, the necessity of ABGSS for social theory will be argued, provided both downward and upward theories are included according to the schema below:

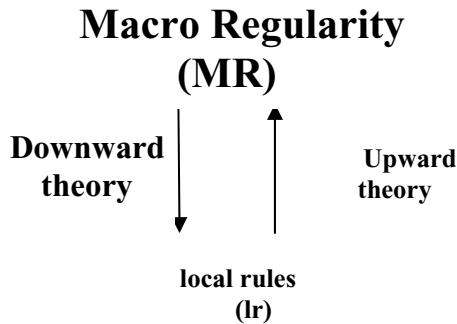


Fig. 1. ABGSS

where upward theory concerns the process from local rules at one level of explanation (individuals), in one direction (bottom-up), to a theory of social processes and properties, and downward theory concerns the process leading from social effects back to local rules.

² Indeed, the semantics of “minimal conditions” ought to be clarified: are they to be meant as simplest possible local rules, or most general? Unlike the latter, the former semantics of minimal conditions include *ad hoc* models.

2 Re-discuss ABGSS

For Epstein, to generate a social phenomenon by means of agent-based simulation requires one to

“situate an initial population of autonomous heterogeneous agents (see also Arthur) in a relevant special environment; allow them to interact according to *simple* local rules, and thereby generate - or ‘grow’ - the macroscopic regularity from the bottom up” (Epstein 1999, 41; *italics are mine*).

Fully to appreciate the heuristic value of ABGSS, one needs to clarify what is meant by explanation. We will turn to this task below.

2.1 Generation and Causal Explanation

This is not the forum for a survey on the philosophical and epistemological debate on explanation and causal typologies. Far too many notions and types of explanation have been identified in the last two thousand years or so to provide an exhaustive survey of the subject matter. However, the idea that explaining phenomena has something to do with generating them is not new. Explanation (cf. Hall³ 2004; but see also Gruene-Yanoff 2006) is often grounded on different types of causes, a subset of which goes back to Hume and his notion of producing causes. What is and how can we tell a producing cause?

“I find in the first place, that whatever objects are considered as causes or effects are *contiguous*; and that nothing can operate in a time or place, which is ever so little removed from those of its existence. Two distant objects may sometimes *seem productive* of each other, they are commonly found upon examination to be *linked by a chain of causes, which are contiguous among themselves, and to the distant objects*” (Hume, 1739, 75)

For the purpose of the present discussion, what is interesting about this definition is the *procedural* nature of explanation: for the British philosopher, to explain a given event means to bridge the gap from producing causes to resulting effects, unfolding the “linked chain of causes” in between. Hence, the process called for by the philosopher is a sort of reverse engineering: reconstruct the whole chain from remote causes to observed effect. But where to start, or better, how far back should one go to avoid ad hoc explanation? I will suggest that producing causes and their link to effects must be hypothesized *independent of generation*: rather than wondering “which are the sufficient conditions to generate a given effect?”, the scientist should ask herself what is a general, convincing explanation, and only afterwards, she should translate it into a generative explanation.

To see why, I turn the reader to an interesting discussion of generative simulation in Gruene-Yanoff (2006). He seems to have an analogous intuition when he observes

³ For Hall, explanation is either based on counterfactual dependence – i.e. explanandum being removed by removing explanans – or on a producing cause.

that producing events are insufficient for causal explanation unless they are shown to be *typical* or *regular causes* by means of “evidential support from independent sources” (ibid.), such as direct observation, well-confirmed theory, or transferable results from experiments. A producing cause is necessary and sufficient for what the author calls a “constitutional” explanation of the produced event, an expression that in a successive rewriting of the paper he replaced with that, far more ambiguous, of “functional” explanation. A functional explanation is usually intended to explain phenomena in terms of the effects they bring about and from which they are reproduced (as Aristotle’s “final cause”), a subset of which are the goals agents pursue, the final causes of their actions. Instead, a “constitutional” explanation, if I understand correctly the point Gruene-Yanoff tries to make, is a form of reverse engineering, reconstructing the process to a given effect, much as Hume suggests.

For Gruene-Yanoff, a generative simulation is neither sufficient for *causal* explanation (which is something Epstein would agree with), unless there is independent evidential support, nor necessary for causal explanation (which instead is something Epstein would not agree with), essentially because there is a lot of theories that provide causal explanation without running on a computer (Simmel effect, Witness effect, etc.), but it is both necessary and sufficient for *constitutional* explanation.

Plenty of reasons for discarding the strong variant of ABGSS are now available, so I will not address the argument of insufficiency any further⁴. Instead, the argument made by Gruene-Yanoff on necessity, and in particular his point about regularity deserves attention.

Indeed, many interesting social theories never received a generative explanation. This is true not only for *analytical* theories, such as game theory, which can easily be translated into programs running on computers (what is the case with simulation-based studies of the evolution of cooperation, as in the immense literature that took impulse from Axelrod’s work in the early 80s). Nor is this the case only with highly *dynamic* theories, such as Simmel’s theory of fashion, which in fact has recently been translated into simulation models (cf. Pedone and Conte, 2000). It is true also of theories like the Witness Effect (WE, cf. Latané and Darley, 1962), which apparently describes no large-scale, dynamic phenomena resulting from heterogeneous agents in interaction. Let us turn our attention to the WE. Is it a good explanation? What would a generative variant of this theory be like? And what would the value added of such a generative variant be?

The WE occurs in social emergencies: whenever bystanders reach and overcome number three, the probability of intervention to the victim’s help has been found (Latané and Darley, 1970) to drop dramatically. Indeed, the probability of a stalemate increases with the number of bystander. Why?

Latané and Darley account for the WE in terms of a *majority rule*, on the grounds of which agents monitor one another and receive inputs as to how interpret and react

⁴ However, it should probably be observed that G-Y argument is weaker than Epstein’s: although one might concede that current simulation-based theories of social phenomena are poorly verified cross-methodologically, there is no reason why in principle this should not happen, and indeed it starts to happen as this type of research is on the increase (see also Di Tosto et al. 2007; Paolucci et al., 2006). Instead, in principle there is no way to ascertain that what the program simulates is a necessary way to get to a given result. Hence, as Epstein states, one given generative explanation can never be said to be necessary.

to (social) events. As three is the minimum number required for a majority to exist, it is also a critical threshold in the occurrence of the WE. This theory, very simple and elegant, has received a great number of evidential support from both experiments and observations, as well as theory-based confirmations (see the psycho-social literature on the influence of majority).

Making an instructive exercise, let us ask ourselves how translate such a theory into an agent-based simulation model. The answer may be easy: implement the majority rule as a *simple* local rule, and look at its effects.

Why does such a use of simulation appear inexorably boring? There is more than one reason.

First, this type of generative simulation would be based upon a previous causal explanation, namely Latané and Darley's theory. Rather than providing an explanation of its own, the simulation shows whether an existing one actually works, producing the expected effects. The generative simulation is therefore used as a proxy of the analytical model, once a causal explanation is available.

Secondly, and moreover, whereas the theory by Latané and Darley has the heuristic and innovative power of a good scientific explanation, its above-described simulation variant would be completely ad hoc. Why, one might ask, the psychosocial theory is explanatory and the simulation model is ad hoc, if in both cases, the same *explanandum* - i.e. WE - is consequent to the same *explanans* - i.e. the majority rule?

Giving it a second thought, the *explanans* is not exactly the same. The psychosocial theory does *not* simply have the majority rule producing WE. Furthermore, according to this theory agents have a majority rule somehow operating in their minds! This is no hair-splitting: in looking for a causal theory, scientists do not content themselves with *any* producing factor. They seek an *informative* explanation, which incorporates *additional understanding of the level of reality* that the phenomena of study belong to. In our example, an explanation that adds understanding on social individuals. That a majority rule leads to WE under specified conditions tells us nothing new about agents' behaviours. Instead, that agents are governed by an internal majority rule, and consequently may interfere negatively with one another under specified conditions, is interesting news. This we learned from the work of Latané and Darley, *independent of* generative simulation.

Let us go back to Hume and his procedural characterization of generative explanation, according to which the causes of any given event are found out to the extent that one has reconstructed the whole chain leading to the event being brought about. This type of explanation is feasible only if a sufficiently informative cause has already been singled out, i.e. if a theory already exists! *Per se*, the generative explanation does not tell you where one should stop in the reverse engineering starting from the event to be explained. Which producing event is sufficiently informative to provide a causal explanation? To state it with Hartmann: "There is no understanding of a process without a detailed understanding of the *individual* contributions to the dynamic model. Curve fitting and *adding more ad hoc terms simply doesn't do the job*" (Hartmann, 1996; italics are mine).

2.2 Generate and Reproduce

Consider natural experiments: the behaviours that are observed in laboratory are sometimes explained in the sense required by Gruene-Yanoff. But are they also generated?

In laboratory, independent variables are manipulated to observe their effects on the target phenomenon, and by this means to reproduce it. Most certainly, Epstein and Hume would convene that in laboratory one cannot unfold the whole chain of events from independent variable to observed effect, which is precisely what a generative explanation is supposed to do.

However if generate means find out local rules allowing the phenomenon to occur⁵, then generation is also allowed in laboratory. Certainly, experimental simulation is better than natural experiments at formulating and observing the effect of local rules. But it won't tell us if these are irrelevant, ad hoc, poorly informative etc. Like a classic black box experiment, which tells us little about what went on between the manipulated variables and the effect observed, simulation per se tells us neither how far to proceed in the reverse engineering from the effect backward to find out interesting local rules, nor whether the algorithm produces the "linked chain" from causes to effects, or a simple shortcut! Where is the difference between the two methodologies? They both look for the minimal sufficient conditions for the effect to occur. Is this what is meant by *growing*? This time, we believe that Hume would divorce from Epstein: whereas the Santa Fe simulator defines growing as finding the minimal sufficient conditions for a given *explanandum* to occur, for the British philosopher remote causes are not explanations until all the intermediate factors from remote causes to the effects are reconstructed. And, I would add, unless those causes are not formulated prior to and independent of generation. Indeed, the ability to provide a *generative explanation* is no more inherent to simulation than it was to laboratory experiments. In both cases, what makes the difference is whether the search for local rules and the link from rules to effect are theory-based or not.

2.3 ABGSS Redefined

To be convincingly backed up, any explanation ought to be tested by means of a generative simulation. From the previous discussion, we draw two other lessons:

- Generative explanation requires a theory of the causes from *which* to grow the effect, otherwise the explanation is irrelevant and a hoc.
- Generative explanation requires a theory of the *linked chain* of events from those causes to effects, otherwise there is no generative explanation but mere reproduction of the effect.

Still, to what extent would social properties be explained by a truly generative ABGSS, based on both theories just mentioned? Only to a partial extent, unfortunately. Let us go back to Epstein's definition for a moment:

"situate an initial population of autonomous heterogeneous agents (see also Arthur) in a relevant special environment; allow them to interact according to simple local rules, and thereby generate - or 'grow' - the macroscopic regularity *from the bottom up*" (Epstein, 1999, 41; italics are mine).

⁵ "Ultimately, 'to explain' (...) society (...) is to identify rules of agent behaviour that account for those dynamics" (Dean et al. 1999, 201).

As I will try to show in the rest of the paper, this view of ABGSS prevents it from being applied to account for the downward process from social properties to local rules. But if this is not accomplished no full account of social phenomena is provided, since causal autonomy of social systems is ruled out.

Let us start from a more detailed discussion of the necessity for a complex theory of the upward process, and then I will turn to the theory of the complementary, downward process.

3 Bottom-Up Theory

Two components of a generative theory of the upward process from local rules to macroscopic effects must be provided, a general theory of local (individual) systems, and a theory of the linked chain. The example discussed below is intended to clarify both components.

3.1 Segregation and Violence

The famous segregation model by Schelling (1971) is usually interpreted as a visual metaphor for social, even ethnic segregation. He discovered that by placing pennies and dimes on a chess-board - where each square represents a house or a lot in a city and pennies and dimes any two groups in society - he could represent the interplay among residential preferences of neighbors. He let agents move around according to various rules. One obvious rule is simply stick to own current location if happy with own neighbors and move to another, or possibly just exit the board entirely, if unhappy.

The author found that a strongly segregated pattern emerges when agents are unhappy with other-type neighbors. Surprisingly, however, this was also the case even with agents expressing a mild preference for own-type neighbors⁶. As any other good Cellular Automata model, it shows a number of interesting things, namely that equilibrium states are not the norm in social phenomena (see later on in the paper); therefore simulations, which do not have this bias, are particularly apt to investigate them, not to mention that simulations are very good at incorporating time and space.

Now, consider the following passage, quoted from another site where a re-elaboration of the Schelling's model is presented: "*rather than a full understanding of the highly complex outcomes of processes, this type of simulations allows us to understand the decision rules of a small number of individual actors*" (quoted from: <http://web.mit.edu/rajsingh/www/lab/alife/schelling.html>; italics are mine). The interest in this passage is twofold: first, it explicitly gives up the objective of a full understanding of the process at hand; secondly, and more interestingly, it states that this type of simulations allows us to *understand* individuals' decision rules.

The latter assertion is unfortunately too optimistic. Indeed, the segregation model itself is much less ambitious than this particular supporter believes to be the case: rather than an understanding of individual decisions, it shows the macro-social effects of individual decisions. In a sense, the model shows that social dynamics, implemented on individual decisions, contributes decisively to social results; there is no

⁶ The reader is invited to specify different happiness rules and check their effects on a demo board, by visiting the following site:
<http://www.econ.iastate.edu/tesfatsi/demos/schelling/schellhp.htm#intro>

need, in sum, for deeply racist individual deciders to obtain a segregated society: such an undesirable effect is produced by members with a mild preference for their own kind. Societies can be more segregated than their members want them to be. This is interesting news. But what does it tell us about individual deciders? Not much.

First of all, and obviously, one might obtain the same result by implementing residential rules that have nothing to do with racist attitudes, for example a preference for homogenous (even if different from own) environments over mixed ones, or characterizing the two subpopulations with different, but internally consistent, attitudes to migrate: this might be interpreted as a metaphor for a different purchasing capacity or a different mobility inherent to different social groups, but has nothing to do with racist attitudes. Precisely because segregation is an emergent effect of *different* local rules, it says nothing about these. Hence, it says nothing about the cause of segregation: no bottom-up theory is available.

Secondly, and more interestingly, simulation *per se* says little about the emergent process either, i.e. the dynamics leading from decisions to the effects. How is this possible, one could ask, if one can stop the simulation at any given tick and see how things are proceeding? Unless one has a theory of intervening factors, what one sees in the simulation is only the consequence, so to speak, of local rules. For example, how complex is the emergent process leading to segregation? To what extent does it interfere, if it does so, with local rules? We will come later to this point.

In the meantime, let us look into some detail one important social phenomenon, which apparently is tightly linked to segregation, i.e. violent crime. It is largely recognized that 95% of it is ethnically homogeneous. According to Kelly (2000), “Violent crime is better explained by *urban flight* than inequality”. What does he mean? Suppose one decides to turn to simulation for an explanation. By means of simulation, one might *generate* the phenomenon with *simple ad hoc* rules of the type “rob out-groups” and “kill in-groups”, but I am afraid one would not advance far deep into understanding the phenomenon of interest.

To say the truth, the segregation model might help find an answer: if neighbours are homogeneous, people will kill in-groups more often than out-groups.

True. But how do we explain that people rob out-groups? Rather than another set of ad hoc rules, of the sort “kill neighbours” and “rob non-neighbours”, another causal theory is needed, about the difference between crimes against person and crimes against property, and the respective underlying motivations. Rather trivially, one might say that, unlike property crime, violence against the person requires no social differences. In addition, and consequently, violence is frequently unplanned, if not unintentional. Far from migrating elsewhere to perpetrate it, violence and riots blow up here and there among neighbours. Perhaps, a less trivial hypothesis suggests that violence against persons derives from competition over the same (scarce) resources, what is often the case among equals and is frequent in poor homogenous neighbours, which probably represent the majority. A still less trivial hypothesis would propose that violence against the person is the consequence of social desegregation, loss of self-esteem and self-derogatory attitudes, and therefore is directed against same-type agents.

Now, one or the other of these hypotheses is needed in order to understand the rules of *individual deciders*: a causal theory of homogeneity in violent crime is needed in order to and before building up a non-trivial simulation of the phenomenon under observation.

In sum, ABGSS needs upward theory to answer two questions prior to and relatively independent of simulation:

- Which local rules? There are three possible answers
 - Simple* rules: these cannot be an acceptable answer because they ad hoc rules are included
 - Plausible* rules (see Epstein 2005): this is too vague and again includes ad hoc rules
 - Suggested by a *general* explicit theory of individual interacting systems. For the theory not to be ad hoc, two properties need be attributed to such systems: they must be autonomous (otherwise the macro-regularity will be hardwired), but liable to downward causation⁷ (on this last property, we will come back in the following section).
- Which process from local rules to macroscopic effect? As the violent crime example shows, although segregation is sufficient to generate the effect, several hypotheses might be brought about as to the intermediate factors (corresponding to Hume's *contiguous* causes). Hence, segregation is not only one among many possible explanations, but also an incomplete one. Can a generative explanation be incomplete? If we consented, we would be bound to accept traditional experimental science as providing generative explanation. Where would the novelty of ABGSS be, then. Hence, rather than accept as an explanation what is sufficient to obtain the macroscopic effect, we accept as generative only a theory-driven, as complete as possible, explanation of the linked chain from cause to effect.

4 Top-Down Theory

In fig. 2, we have a representation of Epstein's argument of ABGSS insufficiency:

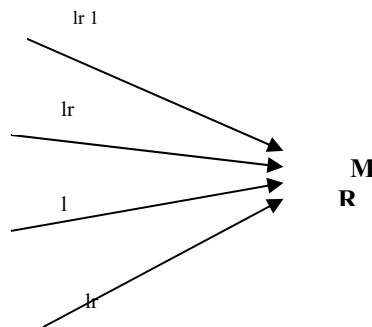


Fig. 2. ABGSS insufficiency

Indeed, this picture is a representation of the argument of *multiple realizations*, according to which a higher-level property is realized by multiple local rules, and vice versa the same local rule may give rise to multiple macroscopic effects.

⁷ The combination of these two properties, indeed, gives rise to what is called social intelligent agents (see Conte 1999).

This argument has led Sawyer (2003) to state that social properties cannot be explained in terms of lower-level rules or mechanisms, which instead, in Epstein's view, is what ABGSS is supposed to do. Who is right?

4.1 Multiple Realizability

Let us consider Sawyer's argument about multiple realizability:

"For example, properties of groups like 'being a church' or 'having an argument' can be realized by a wide range of organizational structures, cultural practices, interactional patterns, and individual beliefs and dispositions."

One given social phenomenon is realized by means of several different local entities. Whether all of them can be described on the grounds of the same mechanism is an option that Sawyer finds empirically irrelevant, if ever possible:

"For example, (...) one can identify a mechanism for the social property "having an argument" such that all possible realizations of this property can be described in systemic terms. If each token instance of the social property has a different mechanistic realization, then the mechanism explaining any one token instance would not be a complete explanation of the property. (...) Sawyer (2002b) showed that a social property might be multiply realized in different systems that are not meaningfully related, using Fodor's (1974) term *wild disjunction* for such higher-level properties. 'Having an argument' or 'being a church' could be realized on one token instance by one mechanism and realized on another occasion by a radically different type of mechanism. If the different realizing mechanisms are not meaningfully related, then the mechanistic, methodologically individualist approach would have limited explanatory power for that social property; an identification of the mechanism on one occasion would indeed provide an explanation of how that social property emerged on that occasion, but the explanation would not extend to the other realizing mechanisms container within that same social property." (p. 268-9)

Hence, social properties are said not to be reducible to the same underlying mechanism, because they are realized on unrelated (wildly disjunctive) multiple realizations of individuals and their interactions.

Sawyer's finds that the property of wildly disjunctive multiple *realizations* of one social property on diverse and semantically unrelated producing mechanisms strongly limits the explanatory power of local mechanisms. Epstein instead believes that this property only shows that such explanatory power is insufficient but necessary.

In the following, I will suggest that Sawyer is right in arguing that such an explanation is fairly limited. However, as I see it, the problem does *not* reside in the *correspondence* between properties at one level and mechanisms at the other: following Hale (1996) and Janich and Psarros (1998), I can easily accept that higher level properties are ontologically dependent on lower level mechanisms. However, this by no means prevents higher-level properties having causal power on properties and

mechanisms at the lower level. But since the view of generative explanation discussed so far is bottom-up, it will never account for this property at the higher level. More specifically:

- Although higher-level (social) entities and properties are realized (implemented) on lower-level mechanisms, nonetheless
- They can have causal power and, consequently, must be brought into play in both the causal and generative explanation of lower-level phenomena, and that this is the case with social effects of individual rules.
- They impact on lower-level entities (say, individuals), thereby modifying the mechanisms on which they are implemented. (If this is what boils down at the bottom of the wild disjunctive argument, so much the better.)

In sum, the problem with ABGSS is not correspondence between levels, but downward causation from social properties to local rules. To see why, we will examine the argument of correspondence between mind and brain.

Sawyer draws upon a well-known debate that is taking place between epistemologists and philosophers of the mind since the mid-70s. Although the debate goes back to the first formulation of Putnam's computational functionalism, it received impulse from the first appearance of a famous work by Jerry Fodor, titled *Special Science. The Disunity of science*. In this work, Fodor claims that the mind is independent of the brain because a mental state is implemented upon a variety of brain states or patterns and vice versa. The consequence the author derives from this premise is that the two levels, the mind and the brain, cannot collapse on one level only. Indeed, to separate the mind from the brain is possible and necessary to understand common (universal) aspects of the mind independent of substrate.

4.2 The Size of the Grain

Such a thesis has undergone several vicissitudes. Denied and restated more than once, at the end of the nineties it encountered a *knock-down argument* expressed by two neuroscientists, Bechtel and Mundale (1997). In essence, their argument unfolds in two steps:

1. While connecting mental states (for ex. hunger) to brain areas in different species, individuals or moments in time in the same individual, the size of the grain is different
2. If the size were kept equal, the argument of multiple realizations would fall: how tell that hunger in men is the same as in mice, or that Jules's pain is the same as Jim's?

Analogously, going back to social properties, who tells that *having an argument* implemented on different local rules would be the same?

4.3 Between Reductionism and Multiple Realizations

Between relying on the argument of multiple realizations and resigning to accept a strong form of reductionism, which takes the social sciences as superfluous constructs

entirely reducible to the study of individuals, another possibility exists. This consists of defining social properties and entities as *ontological dependent* on lower levels but endowed with autonomous *causal power*:

- *Ontological dependence*: all higher levels of reality (not only the macro-social) are ontologically dependent on the lower levels, in that they are implemented on lower levels until the very lowest: higher levels are not ontologically autonomous, since each level of reality is implemented or realized on top of the lower-level one, until the (so far) lowest, i.e. the physical particles’.
- *Causal autonomy*: at any given level of reality, entities may have the power to cause effects at the other levels. Hence, they are necessary for explaining the occurrence of these effects.

Causal autonomy does not require ontological autonomy: to say that a higher level property is implemented on a lower-level mechanism, even a specific one, does not mean to deny causal autonomy to the higher levels. Even if there is no multiple realizations between a given organism and its genes, still no-one would deny causal autonomy to the former on the replication of the latter, nor epistemic necessity in explaining genes’ reproduction.

The necessity to distinguish different levels of reality does not depend on their ontological autonomy, i.e. on their having a “life of their own”. On the other hand, a one-to-one correspondence between lower level realizations and higher level properties should not lead to *eliminativism* (strong reductionism). Think of the big fuss that has been made with regard to mental causation: disturbing Descartes and the epistolary that he shared with the Queen of Poland, some authors (Kim, 2005) have re-proposed the problem whether a mental state can cause or not a behavioural movement. How is that possible? Obviously thanks to the causal process leading from the mental state to the behavioural terminals through a long chain of intermediates, such as synaptic connections, etc. With a somewhat incomprehensible syllogism, from the premise that if it is true, as it is, that any mental state activates a given brain pattern, Kim jumps to the conclusion that there is no need to distinguish one from the other. Such a conclusion is unwarranted at least in the case of proactive action (more on this point, later on in this paper). A mental state is no epiphenomenon of the brain, but a partial and prior cause of action. Analogously, a sequence of instructions running on a computer is not a way to describe the computer, but a cause of the events taking place on the screen when the program is run. It is an intermediate cause of its observed behaviour.

4.4 Weak Thesis and Social Properties

From this digression, we get confirming evidence of what Nagel meant when he said that diverse levels of reality are implemented on one another and reducible to one another through bridge-laws, without denying autonomy and epistemic necessity to any of them. If reductionism asserts only ontological dependence, we can but embrace reductionism. If instead, reductionism is meant as denying causal autonomy to higher levels, in particular the existence of bidirectional relationships among levels - which is a typical feature of complex systems - and considering higher level properties as

epiphenomena of the lower level mechanisms, then reductionism is logically unwarranted and scientifically dangerous. To adopt it means to entomb several other disciplines together with the social sciences, including cognitive and neuroscience as well as biology, to give room to physics only - at least until this will resist the attacks of the antimatter!

As to the mind-brain correspondence, let us consider the following example and ask ourselves what happened to the protagonist:

“If Susan had not had the sudden migraine headache, she would not have experienced frightful anxiety (example taken from Kim, 2005)”.

For the sake of simplicity, let us assume one-to-one correspondence between mental and brain states and in particular that migraine is both a mental state (MS_1) coupled with a brain state (BS_1), and analogously that frightful anxiety is characterized by the coupling of another mental state (MS_2) with another brain state (BS_2). Still, there are two different paths leading from one pattern to the other, and two different, equally possible, configurations corresponding to two different interpretations of the above story.

In story 1, Susan’s migraine activates a brain state and the corresponding mental state that she experienced before a tumour was surgically removed from her brain. Hence, she is scared to death. The link between migraine and anxiety is, so to speak, from the bottom-up: activation spreads through synaptic connections and linked brain states are influenced from one another. Consequently, given the premised one-to-one correspondence, corresponding mental states are brought about (see fig. 2).

Now, consider story 2. Suppose migraine ($MS_1 + BS_1$) is something Susan is not used to. Not knowing how to interpret it, the protagonist will be led to consider it as a novel symptom, perhaps the indicator of a serious disease and this will lead to a state of frightful anxiety ($MS_2 + BS_2$). The link between the two mental/neuronal complex states in story 1 is via mental reasoning and inference (see fig. 3).

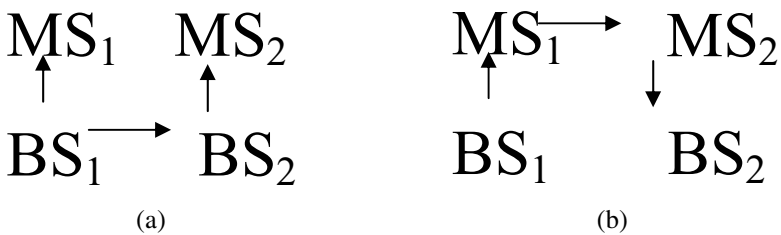


Fig. 3. (a) Brain states activate mental states through activated brain states. (b) Mental states activate brain states by activating other mental states.

How to account for the two stories if MSs collapse on BSs? Independent on one-to-one correspondence Vs multiple realizations, each level may have causal autonomy and determine effect on the other. Hence, both levels are needed to give different explanations for the same effect. *Individual* rules are to social processes what *brain* patterns are to mental processes: both generate higher level properties but are insufficient to explain them.

Going back again to the generative paradigm, the problem with ABGSS is neither insufficiency or, which is the same, multiple realizations. The problem is how account for causal autonomy of social properties. Can ABGSS give such an account?

In his most recent treatment of the generative paradigm, Epstein (2007) explicitly mentions micro-macro mapping as one of the advantages of agent-based computational modeling for behavioural research. Nonetheless, even there, he describes generative explanation as a bottom-up process. According to such a definition, therefore, a generative paradigm fails to account for a fundamental aspect of social theory, namely the way down of the micro-macro process, or, which is another way to put it, downward causation (see fig. 4).

Hence, Sawyer has reason to worry about reductionism in ABGSS not so much because macro-social properties are multiple unrelated realizations of the same local rules, but rather because the impact of social properties on local rules is not an object of study, at least according to Epstein's definition of AGBSS.

In the remaining part of this paper, I will endeavour to show that not only a theory of downward causation is crucial for social theory, but also that it is essential for finding *plausible* (read, scientifically relevant and informative) local rules on which to construct AGBSS. Emergent properties of social phenomena and entities retroact on the agents producing them, and determine second-order emergent effects (Dennett 1995; Gilbert 2001): they start to be perceived by the agents involved, modifying their beliefs and, by this means, reinforcing the corresponding behaviors. To go back to the segregation model, what might happen in residential decisions is the interplay between second-order emergence and the initial local rules. In other words, segregation might initially be determined by mild preferences, but these might be reinforced as soon as agents perceive the effect of segregation (see Gilbert 2001). A strongly segregated pattern resulting from second-order emergence might not be due to a strong preference, but to a process of social "learning". Although the simulation helps investigate this alternative explanation, the latter does not come out from the generative explanation, but from a fruitful interplay between an explorative use of simulation and theoretical advances in causal explanation at the evolutionary level (the theory of downward causation).

4.5 Dynamic Agents and Perpetual Novelty

Back in 1995, Axelrod observed that one bad thing about the mainstream game theoretic accounts of social phenomena was the static view of agents, taken as "given"; conversely, one good thing about agent-based modelling and simulation was a dynamic view of individuals, receiving inputs from their environment, and undergoing social influence.

However, not much simulation research has been carried out about the micro-macro link. Individuals are designed as learning entities, gaining ever more accurate information from the effects of their actions or more successful strategies from observing others' behaviours. In substance, reinforcement learning is shaped on the model of evolution, a fitness formula being always implied. This essentially has led to implement agents' capacity to gain more *accurate information*.

But this is only part of the job required by a dynamic model of agent-hood. Agents undergo social influence, come to share the same beliefs and expectations, squeeze into the same practices, and this type of social influence sometimes leads them to

form inaccurate, even wrong beliefs. Agents do not read the fitness formula in the others' looks or behaviours: what leads them to imitate whom? Why do agents imitate others, and how do they select their social models? A theory of intelligent social learning is badly needed (Conte and Paolucci 2001).

Furthermore, entities and properties emerging from agents' interaction sometimes modify their internal mechanisms and rules. Macro-regularities provide not only new inputs to existing rules, but also new rules. How account for downward causation, by means of a generative paradigm as described above? Let us see some examples of the necessity for a dynamic view of agents that is not reduced to learning.

Agents act on the grounds of beliefs about world-states that they contribute to modify, and which will be modified by their actions.

A typical example of this phenomenon is the minority game, a famous metaphor of which is the *El Farol* bar problem; until very recently, the *El Farol* bar was a trendy pub in Santa Fe, where scholars and students intending to spend their time there found it crowded. To be able to enter, they ought first to guess when it was *less* crowded. But how avoid that making the same guess all drop into the bar the very same night and find it still frustratingly crowded?

As this example clearly indicates, agents act not only on their goals and expectations about the future (see Bickhard 2005; Castelfranchi 2005; Lorini and Falcone 2005), but on a representation of the future that will necessarily prove inadequate, since the future will be modified by their actions.

Note that representations play two different roles in the mind: thanks to their beliefs, which usually are representations of the present world, agents act to achieve their goals, which are representations of the future as agents want it to be. Beliefs not only concern current world states, but also future ones: in the attempt to reach their goals, agents have *expectations* about the world, letting these and other beliefs guide their actions. Where do expectations about future world states come from, especially since they refer to world states that are affected by own and others' actions? This has to do with the coordination games Lewis (1969) referred to (not surprisingly, Lewis found that agents achieve convergence on *salient* equilibriums); but in a minority game, agents have the goal *not* to do what others are doing. Suppose a given option (go to the bar on Friday night) is expected to be preferred by the majority, any option that is different from it – any other night in the week – will be chosen provided the goal to avoid the crowd is more important than that of going out on Friday night. If this is not the case, coordination is not achieved, since agents will not be surprised to find the pub crowded on Friday night. If instead the need for a quiet, peaceful environment is stronger, agents will probably distribute themselves on a wide range of alternatives: they will content themselves with stopping at the bar for a beer or two in any other night of the week. No convergence will emerge as the number of alternatives is too high, and agents will find an easy solution to their problem.

Rather than a paradox, this is a problem of coordination in which the shared preference is *not* to converge on one solution. Obviously, the best way to achieve coordination is to *avoid* the convention. Agents will do so under the conditions that their goal to avoid the crowd is more urgent than the goal to follow the convention. But in any case, *whether they follow the convention or not, agents won't learn the solution to the present problem from experience* but from social beliefs: they are aware that regularity (crowd) follows from convention (go out on Friday night).

As Arthur warns us, the schema of the minority game may be found to apply to a wide range of interesting real-world settings, such as the financial market, and crucial decisions, such as investments.

In a simulation-based study of stock market (reported on in Arthur, 2005), investors were given different hypotheses about price determination, and were endowed with the capacity to act on *the most accurate* and *learn by creating new hypotheses and discarding poorly performing ones*.

The market evolves in one of two regimes: if hypotheses change slowly, agents converge on rational expectations; if they change fast, a chaotic dynamics follows. No convergence emerges.

In the simulation, agents learn to converge by forming hypothetical expectations, which will then be confirmed or disconfirmed through experience. But if change is too fast and no convergence emerges, expectations will always be disconfirmed. When possible, convergence is reached on the most accurate expectations. The lesson drawn by Arthur from these and other examples is the property of *perpetual novelty* of complex social systems, which call for a dynamic approach, focussed on *Equilibrium Formation* (EF) rather than selection.

But in the real world, agents may also converge on the *wrong* expectations! In the real life, the alternative is not either accurate information or no regularity. In real stock markets, sudden crises of trust, whether manipulated and inflated on purpose or not, may be based upon volatile information of the most irrational sort (such as gossip). People draw from one another which option to prefer, which convention to follow, which behaviour to conform to, whom to look after, which expectation to fulfil, etc. Rather than guiding actions to achieve independent goals, expectations often tell agents which goals to achieve.

Although Arthur is right in emphasising the perpetual novelty of societies, one should not underestimate the probability that local equilibriums be determined, not so much by reinforcement learning as by complex mechanisms of downward causation: not only agents modify the world by means of actions based on expectations, but also the world, especially the social world produced by agents, modifies their expectations and their goals, the so-called local rules.

If this is the case, then generating effects only from the bottom up - as in Epstein's formulation - is insufficient. Downward causation is a rather complex phenomenon in its turn consisting of three distinct but equally frequent and often complementary alternatives:

- Learning:
 - Agents act on the grounds of the expectations they have about the future,
 - Form expectations from the effects they produce, and sometimes these effects change so fast that agents have no time to get adjusted to them. This is the lesson that we draw from the Santa Fe artificial stock market.
- Social influence:
 - Agents form expectations by observing and interpreting others
 - Expectations will break current equilibriums or produce novel, however fragile, ones.
- Immurgence:
 - Rules for interpreting others and forming expectations are affected by social properties and entities (norms, authorities, leaders, etc.)

Rules of action are affected and modified by social properties and entities: agents' preferences are modelled on others' preferences, or on what they believe to be others' preferences, or finally on what they believe others (especially well-reputed authorities) want them to prefer.

Rather than from the bottom up, the process of EF goes through multiple loops, from micro to macro and from the latter back to micro again. To understand this process, a theory of local rules is needed, and more specifically a theory of how they are modified by the same effects local rules contribute to produce.

5 Summary and Concluding Remarks

Agent based generative simulation (ABGSS) is a powerful tool for constructing and checking social theories. However, generative simulation should not be meant to replace explanatory theory.

In particular, if it dispenses away with upward theory - i.e. a theory about (a) which local rules, hypothesized prior to and independent of simulation, lead to the observed effect, and (b) how far back in the reverse engineering process from effects to its internal or local causes to go – generative simulation may generate effects without explaining them.

On the other hand, if ABGSS is meant to generate macroscopic regularities only from the bottom up, then it will dispense away with downward theory, thereby ignoring a fundamental ingredient of perpetual novelty and equilibrium formation, which is one main features of social complex systems, i.e. the micro-macro loop. To understand this, not only reinforcement learning, but a general theory of downward causation is needed.

To sum up, the present paper is a warning against a view of science in which the formal tool, be it computational or mathematical, precedes and replaces the theory. This is a mistake, which is not done only by means of equations, but also by means of agent based programming.

Usually, simulation is meant in one or more of the following ways:

- simulation = theory, i.e. generate the phenomena to be explained
- simulation = test the theory.

In all of these cases, the theory is but a metaphor that renders either formulas or equations or programs, or their results, “readable”.

I would like to suggest that there is, or should be, a third way to understand simulation, based upon an interdisciplinary synergetic effort. This synergy ought to converge on providing:

- (computational) theories of individual systems, rather than *plausible* local rules, enriched with mechanisms of downward causation, prior and independent of simulation
- *upward* theories about which mechanisms at the individual agents might lead to observed effects, thanks to evidential support, pre-existing theories and theoretical intuition and translate them into simulation models thereby checking their explicitness and generative power

- downward theories of possible loops from effects obtained back to individuals, taking advantage of the mechanisms of downward causation, thereby checking the generative power of the theory in the complementary direction.

It is in this third way of conceptualizing simulation, that the present work is framed. No methodological revolution can spare scientists the task of searching for good theories, good candidates for explanation. Not even simulation. Indeed, *"What can we learn from simulating poorly understood systems?"* (Simon, 1969).

References

1. Arthur, B.: Out-of-equilibrium Economic and Agent-Based Modelling. In: Judd, K., Tesfatsion, L. (eds.) Handbook of computational economics. Agent Based Computational Economics, vol. 2. Elsevier, Amsterdam (2005)
2. Axelrod, R.: A Model of the Emergence of New Political Actors. In: Gilbert, N., Conte, R. (eds.) Artificial Societies: the Computer Simulation of Social Life. University College Press, London (1995)
3. Bechtel, W., Mundale, J.: Multiple Realizability Revisited. In: Proc. Australian Cognitive Science Society (1997)
4. Bickhard, M.H.: Anticipation and Representation. In: AAAI Fall Symposium, From Reactive to Anticipatory Cognitive Embodied Systems, Hyatt Crystal City in Arlington, Virginia, November 3-6 (2005)
5. Castelfranchi, C.: Mind as an Anticipatory Device: For a Theory of Expectations. In: AAAI Fall Symposium, From Reactive to Anticipatory Cognitive Embodied Systems, Hyatt Crystal City in Arlington, Virginia, November 3-6 (2005)
6. Conte, R.: Artificial Social Intelligence: a necessity for agents' systems developments. Knowledge Engineering Review 14(1), 1–10 (1999)
7. Dean, J.S., Gumerman, G.J., Epstein, J., Axtell, R.L., Swedland, A.C., Parker, M.T., MacCarroll, S.: Understanding Anasazi Culture Change Through Agent Based Modeling. In: Kohler, T.A., Gumerman, G.J. (eds.) Dynamics in Human and Primate Societies: Agent Based Modeling of Social and Spatial Processes. Oxford University Press, New York (1999)
8. Dennett, D.: Darwin's Dangerous Idea. Simon & Schuster, New York (1995)
9. Gilbert, N.: Varieties of emergence. In: Proceedings of Agent 2002 Conference: Social agents: ecology, exchange, and evolution, Chicago, pp. 41–46 (2002)
10. Di Tosto, G., Paolucci, M., Conte, R.: Altruism Among Simple and Smart Vampires. International Journal of Cooperative Information Systems (in press)
11. Fontana, M.: Computer Simulation, Mathematics, and Economics. CESMEP, Torino Dip. Economia S. Cognetti de Martiis, Working Paper No. 06/2005
12. Epstein, J.M.: Generative Social Science. Studies in Agent-Based Computational Modeling. Princeton University Press, Princeton (2007)
13. Epstein, J.M., Hammond, R.A.: Non-explanatory equilibria in extremely simple game with (mostly) unattainable fixed points. Complexity 7(4) (2002)
14. Epstein, J.M.: Remarks on the Foundations of Agent-Based Generative Social Science, The Brookings Institute, Washington, DC, CSED Working Paper No. 41 (2005)
15. Epstein, J.M.: Agent-based computational models and generative social science. Complexity 4(5), 41–57 (1999)

16. Gruene-Yanoff, T.: *The Explanatory Potential of Artificial Societies* (2007), <http://philsci-archive.pitt.edu/archive/00003669/>
17. Hale, B.: Absolute Necessities, *Noûs. Supplement: Philosophical Perspectives, Metaphysics* 30(10), 93–117 (1996)
18. Hall, N.: Two Concepts of Causation. In: Collins, J., Hall, N., Paul, L.A. (eds.) *Causation and Counterfactuals*. MIT Press, Cambridge (2004)
19. Harré, R.: Peripheral Vision. *Discursive Psychology and the Boundaries of Sense. Organization Studies* 25(8), 1435–1453 (2004)
20. Hartmann, S.: The world as a process. In: Hegselmann, R., et al. (eds.) *Modelling and Simulation in the Social Sciences from a Philosophy of Science Point of View*, pp. 77–100. Kluwer, Dordrecht (1996)
21. Hume (1739)
22. Janich, P., Psarros, N. (eds.): *The Autonomy of Chemistry. 3rd Erlenneyer-Colloquy for the Philosophy of Chemistry*, Würzburg, Königshausen & Neumann (1998)
23. Kelly, M.: Inequality and Crime. *The Review of Economics and Statistics* 82(4), 530–539 (2000)
24. Kim, J.: *Physicalism, or Something Near Enough*. Princeton University Press, Princeton (2005)
25. Latané, B., Darley, J.M.: *The unresponsive bystander: Why doesn't he help?* Prentice Hall, Englewood Cliffs (1970)
26. Lewis, D.: *Convention: A Philosophical Study* (1969)
27. Lorini, E., Falcone, R.: Modelling Expectations in Cognitive Agents. In: *AAAI Fall Symposium, From Reactive to Anticipatory Cognitive Embodied Systems*, Hyatt Crystal City in Arlington, Virginia, November 3-6 (2005)
28. Marx, K., Engels, F.: *Selected Works*. Lawrence and Wishart, London (1968)
29. Paolucci, M., Conte, R., Di Tosto, G.: A Model of Social Organization and the Evolution of Food Sharing in Vampire Bats. *Adaptive Behavior* 14(3), 223–239 (2006)
30. Pedone, R., Conte, R.: The Simmel Effect: Imitation and Avoidance in Social Hierarchies. In: Moss, S., Davidsson, P. (eds.) *MABS 2000. LNCS*, vol. 1979, pp. 149–156. Springer, Heidelberg (2001)
31. Sawyer, R.K.: The Mechanisms of Emergence. *Philosophy of the Social Sciences* 34(2), 260–282 (2004)
32. Schelling, T.C.: Dynamic Models of Segregation. *Journal of Mathematical Sociology* 1, 143–186 (1971)
33. Simon, H.: *The Sciences of the Artificial*, 3rd edn. MIT Press, Cambridge (1969/1993)

Talking about ABSS: Functional Descriptions of Models*

Scott Moss

Centre for Policy Modelling
Manchester Metropolitan University Business School
scott@cfpm.org

Abstract. Social simulation research lacks a common framework within which to integrate empirical and abstract models. This lack reflects an epistemological divide within the field. In an attempt to span that divide and in hope that it will lead to subsequent work on integrating abstract and empirical agent based social modelling research, I suggest here that a possibly suitable framework would derive from the mathematical notion of a function as a mapping between a well specified domain and a well specified range. The use of the function as an *informal* framework for the discussion of epistemological issues such as prediction, validation and verification is demonstrated as well as its use for structuring controversy about modelling techniques and applications. An example is drawn from the literature on opinion dynamics to explore the latter use.

Keywords: Agent-based social simulation, validation, verification, model space, opinion dynamics.

1 Introduction

How much of the literature on social simulation is about modelling for modellers? Conversely, how much of the literature is directly about the world in which we live – that is, about society? As of the turn of October, 2006, the 20 most frequently downloaded papers in the flagship journal of the field, *JASSS* included a couple of reviews, 13 papers that cited no direct evidence and five that did report and use qualitative and/or numerical evidence. By way of comparison, the leading economics journal *Econometrica* at the same date listed among its 20 most frequently cited papers, seven papers (including the single most frequently cited paper) that reported and used empirical evidence. While the comparison is by no means exhaustive or definitively scientific, it does not indicate an extravagant interest in empirically based, descriptive modelling amongst leading ABSS researchers. At the very least, it is not obvious that social simulation modellers value empirical studies more than conventional and

* Jim Doran offered helpful comments, for which I am immensely grateful, on a previous draft. He bears no responsibility for the resulting amendments to that draft. The point of departure for my thinking about this paper is a previous paper I wrote with my (then) young colleague, Bruce Edmonds, building on his notion of the “volume” of a model [Edmonds and Moss, 1998]. All failures of understanding or sound analysis in the present paper were introduced by me alone.

traditional social scientists. Our models are sometimes more complicated but, even then, simplicity is still seen as a virtue.

Evidently, while there are empirical papers of which we, as a research community, can and should be proud, the bulk of the work in ABSS is abstract. The abstract models are sometimes intended to reflect some aspect of reality and sometimes to prepare the way for a new kind of social theory and sometimes just to explore the characteristics and robustness of modelling approaches. Without shame, I take the view that the only meaningful purpose of social science is to understand society and social processes thereby to inform social and related policy formation. Producing and understanding modelling techniques to support that objective is clearly important. But producing and understanding modelling techniques that have no relationship to society and real social processes is, I submit, mere self-indulgence. To absorb resources for such self-indulgence that might otherwise be used for empirically and socially valuable development is a scandalous waste.

These issues are epistemological in nature. What knowledge do we seek? What do we know? Why do we think that we know it?

If the knowledge we seek is knowledge about models or modelling technique, then it is hard to see that what we know will be of any wider social value. Because a model space (in the sense described in this paper) is narrowly circumscribed, we can explore that space extensively and, as a result, convince ourselves and others interested in that space that we do indeed know a lot about it. The issue is much more difficult when we turn to human societies.

The whole question of the value of modelling for modellers is not often addressed in the ABSS literature. Perhaps this is because we have not developed a basis for that discourse. Perhaps the basis for discourse has not developed because the issue has not been deemed sufficiently important. Whether the absence of discourse is one, the other or some combination, the purpose of this paper is to raise the question of the value of modelling without direct empirical input. In fact, as modellers we are collectively rank amateurs in comparison with economists or physicists who do physics (as distinct from physicists who do social science). To see the difference in sophistication and detail of the models and the techniques for analysing model properties, have a look at the 20 most frequently cited papers in *Econometrica* at www.blackwell-synergy.com/loi/ecta. Comparing *JASSS* with *Econometrica* makes social simulation look primitive. Where ABSS papers in *JASSS* and elsewhere look good (at least to me) in comparison with papers in *Econometrica* and other social science journals is where the ABSS models are evidence driven.

If these observations have any validity, then a key epistemological question for social simulation modellers and users is whether the bulk of ABSS papers that are not empirically driven are likely to inform or even to provide the foundations for empirically driven social simulation models.

The purpose of this paper is to propose a framework within which to couch a discourse on that issue. To achieve that purpose, I will describe the framework and then consider within that framework an example of a discourse about models for modellers and to suggest why, in that case at least, the modelling is not clearly useful in understanding the social processes that are purportedly represented.

2 The Models as Functions

An elementary definition in secondary school mathematics (as least, as I was taught mathematics) is that a function is a mapping of values in a *domain* into values in a *range*. This is essentially the role of a simulation model. We define a set of initial conditions and then run a computer program that produces an *output trajectory*. The set of all sets of initial conditions that the program can process for the production of output trajectories is the model *domain*. The set of all output trajectories that the model can produce is the model *range*.

Much of what follows is so natural as hardly to need stating. It is stated only for purposes of clarity.

The initial conditions of any run of an ABSS model includes a set of agents. Each agent is itself a function. At each time step, it perceives aspects of its environment. It processes those perceptions and then executes some action affecting itself and/or its environment. What the agent can perceive is its domain and the changes it can evoke are its range. The environment of an agent includes the other agents it can perceive. The social network is therefore an element of the model domain determining the network perceived by each agent as part of its individual domain. At each time step, the program maps the prevailing element of the domain into an element in the same space – that is the same collection of potential agents and possible social networks. An output trajectory is a sequence of elements in that same space so that the range of the model is a sequence of points in the space of the domain. In a rulebased system, it is typically possible for rules to define new rules, thereby to add new elements to the range space that were not present in the domain space.

2.1 The Model Space

More than 10 years ago, Bruce Edmonds and I [Edmonds and Moss, 1998] defined the model volume as the Cartesian product of the range and the domain sets (the notion of a model volume was Bruce’s). More recently, Bruce has coined the term “model space” to mean, I think, the same thing. In this paper we shall refer to the Cartesian product of the domain and the range as the model space. In some models, such as those in which agent behaviour is determined by rules it is possible for the model space to expand with passing time steps. Perhaps this is a general characteristic of models in which agents learn. In such cases, the model space itself is endogenous to the model. The model space is a useful concept because one of the issues to be considered below in section 6 is the “exploration of the model space”.

2.2 The Mapping Function

The mapping function of a model is essentially the programming language and libraries that cause the agents to become active and process their perceptions to arrive at their actions. So the mapping function would include the scheduler in Swarm, RePast and MASON. In models that, for example, incorporate the Jess declarative programming libraries with (say) RePast, the mapping function includes the inference engine, the conflict resolution mechanism, database procedures and so on of Jess as well as the RePast scheduler determining the order in which agents are activated and the code for storing and manipulating data for the model as a whole.

In general, we do not know the formal properties of the mapping function. We do know that programs written in any language, if they do not crash, are sound and consistent with respect to that language and that we can assume that they are decidable if programs complete under a wide range of input configurations (or domain elements). We cannot rely on completeness. The only attempts of which I am aware to develop programming languages that give an element of formal proof to the mapping functions of ABSS models were SDML and DESIRE.

3 Data Generating Mechanisms

Econometricians have defined *data generating mechanism* [Phillips and Perron, 1988, p. 336] to mean a function with lagged values (*e.g.*, an autoregressive process) that, given a set of initial values, will produce by repeated application a stream of successor states. This is equivalent to defining the data generating mechanism as the domain, mapping function and range of a model. A model is, in this sense, an *artificial* DGM. When models are intended to capture aspects of social reality we are in effect supposing that there is an underlying correct DGM. Sometimes, models are intended to capture the perceptions of specific individuals (or shared within a group of individuals). Incorporating those perceptions in a model is intended to inform the artificial DGM as a representation of a perception of the relevant aspects of the social (real) DGM. To maintain the parallel with artificial DGMs, we should say that the social state prevailing at the initial time of a social study is the relevant element of the domain and the time pattern of the later observations is said to be an element of the domain. The specification of the mapping function will depend on the institutional and physical infrastructure. If the target social processes are, for example, the processes that comprise a financial market, then the communications infrastructure and the administrators and clerical staff that process data and publicise prices and volumes *could* comprise the mapping function.

In practice, the choice of elements of society that we treat as the social mapping function naturally depends on the problem being addressed. In the example of the financial markets, it is entirely possible that the informational and administrative arrangements would themselves be part of the subject of interest. In that case, the relevant individual decision makers and those who follow prescribed procedures would be represented in the model domain and range space and would therefore be considered in the initial conditions and therefore the domain of the societal DGM.

4 Prediction

One example of the utility of treating models as functions, is in discussions of prediction. This is not an entirely trivial issue for social simulation whenever we seek to understand broad properties of a system without capturing the same scale or, frequently enough, being able to choose time steps for the model that correspond to the available data. This is particularly important when some data is at a much finer grain, and therefore exhibits different properties, than other data series. This was an issue facing Moss and colleagues [Moss and Edmonds, 2005] in modelling household water

demand. The model with monthly data about precipitation and temperature, calculating ground water values using an algorithm based on month-based parameters, produced domestic water consumption values with episodes of clustered volatility, hence leptokurtosis, that did not show up in monthly water consumption data. However, we recognised that using course grain data (monthly rather and, say, daily data) amounts to taking the average of observations of more fine grain data and, by the central limit theorem, the distribution of averages converges to the normal distribution whether or not the underlying data is normally distributed. Collaborators from the Swedish Environment Institute were able to obtain daily water consumption data for residential neighbourhoods which confirmed the *prediction* that distributions of household water consumption volumes are leptokurtic.

This result is a different, and much less precise, kind of prediction than (say) econometric forecasting where, using monthly or quarterly data, predictions are made regarding future values at a grain of time not less than the minimum time grain of the data. In the water demand model, the prediction was about the presence of leptokurtosis and also the unpredictability of volatile episodes of changes in water consumption volumes. There were no predictions of specific values of consumption volumes. The water demand prediction was, in this sense, a weaker form of prediction than econometric prediction. At least up to the present, we have no means of predicting any the values of any of the moments of the distribution other than to say that the fourth moment (kurtosis) is large relative to the same moment of any distribution drawn from a population with finite variance. As Fama [1963] has pointed out, none of the standard forecasting and goodness-of-fit algorithms are applicable in these conditions.

We therefore distinguish between *strong* and *weak* prediction. Strong prediction is the assertion that the current social state maps into a point in the model domain that itself maps into a specific output trajectory in the model range. Weak prediction is the assertion that the range of a model or some subset of the range constituting the image of some subset of the domain has specified properties such as leptokurtosis.

5 Validation and Verification

Prediction has, of course, conventionally been associated with model validation. Economists in particular, following Friedman [1953], typically assert that theories are validated by their predictions without regard for the realism of their assumptions. In the framework reported here, this amounts to the assertion that

- every element in the domain of the social DGM maps into an element in the domain of the model DGM
- every element in the model domain space maps into an element in the model range space
- every element in the model range space maps into an element in the range of the social DGM

If it is allowed that models can be expected to generate correct predictions only when specified conditions of application are satisfied, then the first item in the list above is modified to state that a subset of the elements in the social DGM maps into the model domain. Moreover, if there is no mapping from the currently prevailing element of the social DGM into the model DGM domain, then no prediction is to be made.

Evidently, this account of prediction and validation turns on the feasibility of strong prediction. When strong prediction is not possible, we have to rely on weak prediction for model validation. In this case,

- a subset of the social DGM domain maps into the model DGM domain
- the range of the model DGM has a set of characteristics that might be statistical (such as leptokurtosis) or might be qualitative or some combination of the two.

Each element of the model DGM domain and range has an associated set of agents with their current states as well as the macro state of the model. It is therefore possible to compare the micro (agent) states with the circumstances and behaviour of real decision makers and also to compare the macro state with more highly aggregated data from the real world. This is what Moss and Edmonds [2005] called *cross-validation*. While cross-validation relies on weak prediction at macro level, it is more confirmatory at micro level. That is, micro validation involves checking with informants and other domain experts that the behaviour of the agents conforms to that of the individuals or organisational units the agents represent.

Verification – the process of formally proving the properties of a model – is not normally an issue in social simulation modelling. However, it is possible that treating models explicitly as functional relations could support some verification. For example, we know from the Brouwer fixed point theorem that any function mapping a domain into itself has at least one element of the domain that maps into that same element. This is uncomfortable for those of us who abhor the notion of a social equilibrium since this is a sufficient condition for an equilibrium to exist. On the other hand, even if there is such an equilibrium configuration, it is a formal possibility that, in social models with real numbered variables, the equilibrium is unrealisable with finite precision arithmetic as inevitably used in computational models. It would be surprising if other formal relations could not be proved for models as functions but, if they can, it will be important to distinguish between such formal results and computationally feasible results. Since decision makers cannot in fact use infinite precision arithmetic in reaching their decisions, such formal but computationally infeasible properties of models can surely be ignored.

6 Exploring the Model Space

Model space exploration frequently takes the form of a sequence of random selections of elements in the model domain and then some graphical or statistical processing of data from the resulting output trajectories. There are also more selective and also formal approaches to model space exploration. We distinguish four such approaches:

- *Random exploration*: Monte Carlo studies and the like.
- *Selective exploration*: Choose elements from a subset of the domain that is claimed to be in some sense realistic or relevant.
- *Formal exploration*: Use logical (including mathematical) means to identify the boundary of the meaningful or feasible subset of the domain and explore the range corresponding to that subset. Terán [2001] demonstrated the feasibility of formal model exploration by exploiting the features of SDML, a

strictly declarative modelling language that was sound and consistent with respect to a fragment of strongly grounded autoepistemic logic.

- *Mixed formal-random exploration*: Choose random values for initial states of a model and use, for example, constraint logic to determine whether it is feasible. If it is feasible, then run the model to determine the corresponding output trajectory. Repeated applications will determine trajectories in the range of the model and others that are not in the range of the model, thereby to determine with increasing fineness the boundaries of both the domain and the range.

7 Example: Opinion Dynamics

In the opinion dynamics literature, opinions are usually represented by real numbers in the unit interval¹. The virtues of the numerical representation of essentially qualitative phenomena such as opinions are (1) the greater efficiency of numerical calculation over the manipulation of symbols and (2) the ability to apply numerical processing techniques, especially those such as mean field theorems that are commonly used by physicists. In restricting values to numerical fields, both the model domain and the model range are within sets of numbers – in the opinion dynamics case, usually the field of real numbers in the unit interval.

There are two questions to be addressed in this regard. The first is whether the results are sensitive to the particular choice of numerical representation. A second question is whether real numbers *can* provide a useful representation of opinions. These questions, which are considered in turn, are essentially questions about the appropriate representation of the social domain and range.

7.1 Deffuant-Edmonds Controversy

Deffuant's most recent version of the opinion dynamics model [Deffuant, 2006] supplements the representation of opinion by real numbers with a probability distribution to represent an agent's certainty about its opinion. The mean of the distribution is the current opinion and uncertainty is indicated by the variance in the usual way [Tobin, 1958]. In addition, each agent can be influenced by other agents. The influence of one agent upon another is represented by a function of the opinion of that other agent. Deffuant tries different functional forms. The agents known to each agent is determined by a social network in which the links represent acquaintanceship and the nodes represent agents. Deffuant explores his model space by running simulations with different probabilistic distributions (formulae for uncertainty), different functions determining inter-agent influence, different social network structures, and some other differences. He determines the regions of his model space where opinions converge to one or to both extremes in the unit interval.

Bruce Edmonds [2006] replicated the structure of earlier versions of the Deffuant model that obtained much the same results on extremism and showed that similar but different functional representations of the strength of influence between agents were more likely to lead to moderation rather than extremism.

¹ Other numerical representations are sometimes used.

Deffuant's [Deffuant, 2006] response was this:

“As suggested in Edmonds (2006), a good way to test the robustness of the results is to check if the addition of some noise changes the results. However, in Edmonds proposal the same noise is applied to all individuals. Therefore, extremists can suddenly strongly change their opinion. This is in contradiction with the assumption of the model stating that the extremists are very convinced of their opinion. To keep the rationale of the model, we consider that the level of random variations in the opinion is moderated by the level of uncertainty.”

On this basis, Deffuant assumed that variance of the probability function tends to be smaller at the extremes than in the middle of the unit interval. With this change in the Edmonds specification, opinion clustering at the extremes of the unit interval was re-established.

7.2 Critique and Response in the DGM Framework

In effect, the domain and range of the Edmonds model differed from that of the extant opinion dynamics models only in the representation of uncertainty of opinion and the functional representation of inter-agent influence. There were no obvious substantive differences in the mapping functions of the two models. As in Fig.1, we can imagine an abstract domain comprised by agents that have numerically valued opinions and functional forms relating other agents' opinions to changes in their own opinions, probability functions to represent uncertainty of opinion for each agent and social network structures. There might be some intersection between the subsets of that abstract domain representing the Edmonds and Deffuant models, respectively, but there is also a substantial difference. The ranges of the respective models are non-intersecting since the output trajectories of the Deffuant model and its predecessors are extreme opinion clusters whilst, in the Edmonds model, the opinion clusters are moderate.

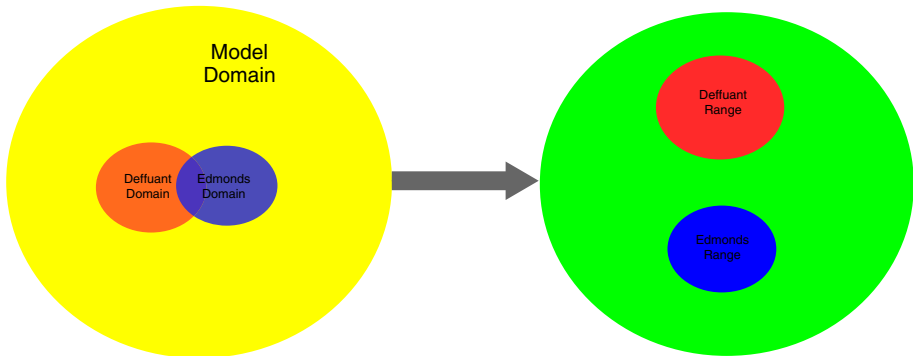


Fig. 1. Domains and ranges implied by Deffuant [2006] and Bruce Edmonds [2006]

Neither Edmonds nor Deffuant suggests any specific mapping from their abstract model domains to any social domain. Indeed, Edmond's concerns are strictly to explore the consequences of using number to represent qualitative phenomena. Essentially, his point was that the results obtained by representing opinions by real numbers

are sensitive to the particular functions that take those numbers as arguments and, at the same time, there is no obvious way to identify any specific, actual social processes and human behaviour represented by the numerical representations and the functions into which they enter. Deffuant is more concerned with the social implications of his model but he is careful not to make any claims that would be interpreted here as a direct mapping between his model space and actual social phenomena. In his words,

“The analysis of these differences (and particularly the differences between the convergence patterns obtained on a lattice and the ones obtained on a random network with 4 neighbours on average) led us to identify several explanatory features of these results. One of them seems particularly noticeable: The single extreme convergence takes place when the moderates remain clustered and isolated as long as possible from the moderates. In this case, the extremists remain influential, without being influenced by the moderates. The clustering of the moderates is favoured by networks where the individuals are all close to each other in the graph.

This observation could be associated to the following interpretation: to remain effective, the extremists must always remain different from the mass of moderates. The big risk for them is to become normalized. This result holds for all the studied model variants.” [Deffuant, 2006]

There are certainly societies in which extremism is dominant or, at least, where the extremists are far more visible than moderates. Northern Ireland through the 1970s and 1980s is a fairly recent and obvious example. There are other societies where extremism appears to be socially marginal and extremists do not constitute large clusters. Either these models are nothing more than explorations of formal relations or they have some empirical content. In the terminology used here, if they have empirical content there will be a mapping between the model domain and range on the one hand and some clear description of the state of societies on the other. Whether such a mapping exists depends on whether there is some unambiguous (if not univalent) mapping between real numbers in the unit interval on the one hand and a natural representation of opinions on the other. If there is such a mapping, then we can take further steps to investigate how to map the other aspects of the model into the real world.

7.3 Opinions as Numbers

The question addressed here is not the technical one raised by Edmonds but rather the question of whether a real number in the unit interval (or in any other number) is a useful analytical representation of the semantics of opinion. In order to investigate this issue, I have looked at opinion research into a topic where extreme views are held with more or less conviction on both sides of an argument: gun ownership and gun control in the USA. Questions and results from a range of surveys conducted over the past 20 years were, at the time of writing, available from the web site of the Roper Center for Public Opinion Research in the University of Connecticut². Those that would allow for different degrees of certainty include this, conducted by the Gallup Organization in December, 1993:

² <http://www.ropercenter.uconn.edu>

QUESTION: Next, I'm going to read you some things people say would occur if there were stricter laws on the buying and owning of guns. For each one, please tell me whether you strongly agree, somewhat agree, somewhat disagree, or strongly disagree.)... Stricter laws would give too much power to the government over average citizens.

ANSWERS:

Strongly agree 30
Somewhat agree 22
Somewhat disagree 26
Strongly disagree 21
Don't know/Refused 1

There were a range of such statements though the Roper Center does not report them or the responses to them.

Another question from a survey by ICR Survey Research Group, in August, 1997 was: *Do you think that the current collection of laws that govern gun ownership and regulation are very adequate, somewhat adequate, not too adequate, or not at all adequate?* with the response distribution:

Very adequate 10
Somewhat adequate 38
Not too adequate 21
Not at all adequate 23
Don't know 7
Refused 1

These surveys were undertaken by different organisations in different years. Both surveys were conducted by telephone interview but the first sample size was about 1000 and the second about 2000. It is possible that a given sample of respondents would feel more strongly – more certain, in Deffuant's terms – about the statement concerning stricter laws than they would about the adequacy of existing laws. But we cannot assert that. The essential point here is that polling organisations ask a range of questions to probe respondents views about different aspects of an issue. Some questions are check questions to ensure that respondents are giving thoughtful answers. But, in general, issues such as gun control are relatively complicated and do not readily, or for some purposes usefully, map into a single real number. In this terminology, we would have to say that there is a real data generating mechanism and a *sui generis* set of data generating mechanisms under the rubric of opinion dynamics. Deffuant and his co-authors describe their data generating mechanisms using words that give the impression that they are related to social processes giving rise to changing distributions of opinions in actual societies though, as we have seen, Deffuant is careful not to state that his is in any sense a *correct* representation of such processes. Edmonds does not address the relationship between his data generating mechanism and societal data generating mechanisms at all. He is simply exploring relationships among domains, mapping functions and ranges, demonstrating that some apparently innocuous modifications in domains have non-trivial consequences for the definition of corresponding ranges.

8 Conclusion

The problem raised in this paper is the lack of discussion about what in general terms constitutes a good or an acceptable agent based formal representation of real target societies and social institutions. The lack of any general framework for such a discussion can hardly be an encouragement for social simulation modellers to address the issues involved. As a first step in providing such a framework, I have suggested that we recognise explicitly that models have the properties (and formally are) functions in the sense of mappings from a domain of agents (which are themselves functions) into a range of agents. Any particular simulation experiment will have a set of initial conditions defined by a particular set of agents and their own parameters and functions including possibly rulebases and databases. Each set of initial conditions will map into a sequence of subsequent states. In discussing the utility of any particular model or set of models, it might be helpful to consider explicitly the ways in which the domains and ranges map into the target institutions and the correspondence between the mapping function of the model and observed social processes.

The example of opinion dynamics was adduced to exemplify the issues. The domain of opinion dynamics issues includes agents have associated numbers said to represent their opinions and a process (mapping) leading from the opinion of each agent at one time step to agent's opinion at the next time step. Edmonds pointed out the sensitivity of the mapping to the functional form and its parameters. Evidence from opinion surveys was cited to indicate that some opinions are apparently multi-faceted and that a single real number is not an obviously descriptive representation of something as complex as an opinion.

The point is not simply to criticise Guillaume Deffuant and his colleagues or to attack the opinion dynamics literature but to create a framework in which such discussions can usefully take place. The value of the framework – whether that suggested here or some other framework – is to enable us to collect issues, disagreements and criticisms in a coherent fashion. I do not believe that we have yet succeeded in combining breadth and coherence in the discussions of our different approaches to agent based social simulation.

References

1. Edmonds, B.: Assessing the safety of (numerical) representation in social simulation. In: Agent-based computational modelling, pp. 195–214. Physica Verlag, Heidelberg (2006)
2. Deffuant, G.: Comparing extremism propagation patterns in continuous opinion models. *Journal of Artificial Societies and Social Simulation* 9(3) (2006)
3. Edmonds, B., Moss, S.: Modelling learning as modelling. *Cybernetics and Systems* 29, 5–37 (1998)
4. Fama, E.F.: Mandelbrot and the stable paretian hypothesis. *Journal of Business* 36(4), 420–429 (1963)
5. Friedman, M.: The methodology of positive economics. In: *Essays on Positive Economics*. University of Chicago Press, Chicago (1953)

6. Moss, S., Edmonds, B.: Sociology and simulation: Statistical and qualitative cross-validation. *American Journal of Sociology* 110(4), 1095–1131 (2005)
7. Phillips, P.C.B., Perron, P.: Testing for a unit root in time series regression. *Biometrika* 75(2), 335–346 (1988)
8. Terán, O.: Emergent Tendencies in Multi-Agent-based Simulations using Constraint-based Methods to Effect Practical Proofs over Finite Subsets of Simulation Outcomes. PhD thesis, Centre for Policy Modelling, Manchester Metropolitan University Business School (2001), <http://cfpm.org/cpmrep86.html>
9. Tobin, J.: Liquidity preference as behavior towards risk. *The Review of Economic Studies* 25(2), 65–86 (1958)

What Does Emergence in Computer Simulations? Simulation between Epistemological and Ontological Emergence

Alex Schmid

Savannah Simulations AG, Alte Dorfstrasse 24
8704 Herrliberg, Switzerland

a.schmid@savannah-simulations.com

Abstract. Emergence is generally considered a fundamental property of complex systems. Being a central but notoriously ill defined notion, concepts of emergence fundamentally oscillate between epistemological and ontological interpretations. The paper relates these philosophical perspectives of emergence to the interpretation of emergence in computer simulation. It concludes that most arguments point to the fact that computer simulation deals with epistemological emergence only. However, there is no conclusive argument that computer simulation in principle is unable to model ontological emergence. Finally, the paper argues for mathematics being a restricted description what concerns all possible emergent levels not yet realised.

Keywords: Computer simulations, emergence, ontological and epistemological emergence.

1 Introduction

An important aspect of computer simulation, especially the simulation of dynamic non-linear complex systems, consists in the analysis of novel emergent phenomena where micro-behaviours generate “unexpected”, seemingly unpredictable macro-behaviours of some kind. Especially in social simulation where a high complexity is given, the investigation of emergent properties or macro-states based on the behaviour of micro-states of systems or agents are one of the main goals of computer simulation. Unfortunately in most cases, not only in respect to computer simulation, the exact meaning of the term “emergence” and its use is rarely specified. Although research and use of the term emergence spans from quantum physics (Laughlin, 2005) to complex biological systems (Kauffman, 2000) and cosmology (Smolin, 1999), there is no agreed taxonomy of emergence neither in science, philosophy of science nor in philosophy in general.

In this paper I will argue in favour of the most promising philosophical distinction of emergence at the moment, namely between *epistemological emergence* and *ontological emergence*, and some conclusions are drawn what regards this interpretation of emergence and computer simulation. The research question can be summarized as follows: given that one or both of these kinds of emergence hold in reality, what kind

of emergence does and can computer simulation represent or emulate? Do computer simulations based on their specific means and properties represent epistemological as well as ontological emergence, or just one of them, and what are the technical and philosophical or epistemological reasons for this specific kind of representation?

The question is not only classificatory, but of fundamental importance for the epistemological and scientific standing of computer simulation. The central question can be summarized as follows: By conducting simulation studies, do computers just allow to calculate and mimic the vast complexity of systems and their emergent properties that can be reduced in principal to simpler parts and behaviours (i.e. epistemological emergence) or are we able to reproduce by computation the true genesis of new, un-reducible levels of reality (i.e. ontological emergence)?

2 Emergence and Part-Whole Reductionism

The most general definition of emergence states that a part-whole reductionism does not apply. Part-whole reductionism says that all system (whole) properties can be deduced from and reduced to their constituent and basic parts. In turn, what characterizes emergent properties is that they cannot (or only partially) be deduced – and explained - from the properties or interactions of the parts.

Part-whole reductionism belongs to the cornerstones of science, as Scharf (1989) states: “*the program for the unity of science is a program for universal micro-reduction*”. Also the most widely accepted model of scientific explanation, the deductive-nomological model, is largely inspired by this universal micro-reduction statement.

Today, the most fruitful classification of different types of emergence seems to be the differentiation between “epistemological” and “ontological” emergence. I will briefly resume this difference. Epistemological emergence can be defined as follows:

“A property of an object or system is epistemologically emergent if the property is reducible to or determined by the intrinsic properties of the ultimate constituents of the object or system, while at the same time it is very difficult for us to explain, predict or derive the property on the basis of the ultimate constituents. Epistemologically emergent properties are novel only at the level of description” (Silberstein, McGeever, 1999).

Consequently, ontological emergent features are features of systems or wholes that “possess causal capacities not reducible to any of the intrinsic causal capacities of the parts nor to any of the (reducible) relations between them” (Silberstein, McGeever, 1999). It is important here to hold in mind the distinction of real “causal capacities”, connections or actions versus logical causality or abstract relations. We will see that the former is a distinctive feature of the description of ontological emergence, the latter of epistemological *emergence*. In addition, *synergism* (combined or cooperative effects between objects or systems), *novelty*, *irreducibility*, *unpredictability*, *coherence* and *historicity* are further prominent properties of emergence in literature (Achinov/Fuchs, 2003). Of these concepts I will use only those that are important for the topic in question. In addition to the distinct properties of emergence, there always has to be considered the *objects of emergence*. The question is what is specifically emerging, *properties*, *entities*, *new laws* or *dynamics* or other sorts of emerging phenomena (Silberstein/Geever, 1999)?

Thinking about the reality and distribution of epistemological and ontological emergence in the universe, we can find three different propositions that might describe a solution:

- 1) *reality contains epistemological as well as ontological emergence,*
- 2) *reality contains either epistemological or ontological emergence, but not both, and*
- 3) *neither epistemological nor ontological emergence are true in this world, but another description or process we don't know yet.*

Obviously, proposition 3) is the most unlikely, since if 3) would be true, all scientific knowledge up to this day would be nothing but some sort of illusion. Silberstein and McGeever give a further account of 1) by stating: “*Most cases of emergence are epistemological.*” (Silberstein/McGeever, 1999) Most emerging properties in the universe are reducible to the properties and behaviours of their parts, but there remains the question if ontological emergence exists at all. Having a closer look at 1), Silberstein and McGeever conclude that epistemological emergence logically cannot entail ontological emergence, because it is defined to preclude it. If something is reducible in principle it cannot be irreducible at the same time. Still, this argument is consistent with 1) postulating that epistemological emergence can co-exists with ontological emergence.

2) seems to be a solution of 1) insofar we would find out that 1) is wrong, we would be committed to 2) – excluding 3) as the most unlikely – namely that reality contains either epistemological or ontological emergence.

According to Silberstein and McGeever it is also not necessarily true that ontological emergence entails epistemological emergence. This would only be the case if ontological emergence would presuppose epistemological emergence, either logically, causally or otherwise. This means that it could be that ontological emergence is something “on top and above” epistemological emergence whatever specific relation that would be.

From the standpoint of philosophy of science it is obvious that the more interesting question regarding emergence deals with ontological emergence that directly opposes scientific reductionism, whereas epistemological emergence follows the scientific paradigm. Therefore I will state the central argument of the relation of computer simulation and emergence as follows:

Given the possibility that reality exhibits to some extent ontological emergence, what is the exact relation of computer simulation to ontological emergence? Is computer simulation capable to simulate and reconstruct ontological emergence? And if not, what exactly makes it impossible for computer simulation to emulate or elicit ontological emergence?

3 Mathematics, Computation and Emergence

To put the conclusion first: It is very likely that computer simulation mainly has to do with epistemological emergence, following the aforementioned statement by Silberstein and McGeever.

Computer simulation is basically the mathematical and logical reconstruction and dynamic approximation of the complex properties, dynamics and relations of the components of a natural, social or artificial system, coded as a computer program. For our purpose, the specific simulation technique, be it difference equations, cellular automata, agent-based modelling, genetic algorithm or any other modelling approach, is not important. We concentrate only on the most general features of the “description mode” of computer simulation and these are mathematics and logic. Of course, mathematics includes logic, but we separate logic to point out semantic, logical and language-based simulations for example in artificial intelligence that don’t rely exclusively on mathematics. After all, *mathematics, natural and formal languages* are the main “tools” we have for modelling natural or social systems.

To take up an epistemological and mathematical view on simulation inevitably brings forth considerations upon different concepts in philosophy of mathematics, departing from mathematical realism and empiricism to constructivism, fictionalism or social constructivism (Shapiro, 2000). Do we have to analyse all these concepts to evaluate the relation of mathematics to epistemological and ontological emergence?

The fundamental point or divergence in philosophy of mathematics is the question if mathematical entities – numbers, axioms etc. – depend on the existence of human mind or if they are independent entities. Are they objective facts or subjective constructs? Do there exist mathematical entities in the universe even if there would have never been any human mind (objective view of mathematics) or is mathematics a (social) construct of human mind that is subject to revision like any other empirical endeavour of human kind (subjective view of mathematics)? All theories in philosophy of mathematics tend to one or the other side, although most philosophers of mathematics today would not commit themselves solely to one of the extreme positions. It seems that mathematics involves objective as well as subjective/constructive properties.

For our purpose, we don’t have to be committed to any specific interpretation in philosophy of mathematics concerning the relationship between mathematics and reality. It is enough to assume two propositions:

- 1) *it is true that mathematics has at least one common property with reality*
- 2) *it is true that mathematics cannot have all the properties that constitute reality.*

In other words, 1) mathematics has some relation with reality and it is impossible that it has an “unconnected” existence separate from causalities of reality, 2) it is a fundamental property of mathematics to be abstract since the properties of the two “worlds” cannot be identical. The fundamental abstraction implied in the concept of a “model” is another way to state 2). The specific property of this minimal relation, be it representational, descriptive or any other property, is not important. For example, if we assume that ontological emergence exists, then by 1) we can conclude that mathematics in principle is capable of participating in some way in ontological emergence, and by 2) that it cannot “be” in an existential way ontological emergence.

Bertuglia and Vaio (2005) state this fact as follows:

“Mathematics...can be defined as the art of creating models, extremely abstract and simplified models, models, so to speak, in black and white, that describe the deepest essence, the skeleton (or, rather, what seems to us to be such) of a real situation.”

I would not agree with the term “essence” used in this statement, but it describes the true abstract nature of mathematics that will never be ontological equivalent to reality.

Having defined mathematics for our purpose, there remains the task to analyse the epistemological status of computer programs in relation to emergence. Thus the question: does the further coding of a mathematical model add anything decisive to the aforementioned argument? From an epistemic perspective I would argue that computer coding adds nothing fundamentally new to the aforementioned argument. 1) and 2) also apply for coding although one might suspect that there is a gradual difference between mathematical and coded models.

4 Simulation and Emergence

If we assume that ontological emergence exists, how can we further analyse the relation between simulation and given ontological emergence, having examined properties of mathematics and emergence? It seems obvious that the computed emergent properties of a simulation model must be reducible or must be able to be derived “in principle” from the underlying mathematical structure. Therefore, it seems that the encoded laws and relations of a mathematical model can’t generate in principle new laws that are not reducible to the underlying laws. Theoretically, if these mathematical or encoded laws and regularities have any similarity with real facts and processes then the “in principle” conclusion is wrong (according to proposition 1), since these “real” facts and processes are capable in nature to generate this non-reducible ontological emergent properties, and mathematics can in principle participate in this genesis – even if this means to capture the “one and only” ontological property which would possibly have an astronomically small probability to happen.

Generally, there are at least three possible arguments concerning the implication of emergence in computer simulations:

1) *Simulation can only restate or emulate epistemological emergence* - it is the only goal and capability of simulation to overcome the “difficulty” of (analytic) description through the numerical solutions of the complex interactions and relations of the parts. Simulated or computed emergence remains therefore always on the level of epistemological emergence: the emergent ant hill has no emergent ontological properties, in fact, we are only incapable to analytically describe the emergence of the hill through interactions of thousands of ants and their behaviours. Simulation “resolves” this descriptive gap and adds a new descriptive (and reductive) way to the understanding and calculation of such utterly complex phenomena. There might be some ontological emergence “out there”, but there is no way to capture the evolution and the dynamics of these new levels using computer simulation.

2) *Given that simulation restates epistemological emergence only, the failure of epistemological explanation or simulation is still a “negative prove” of ontological emergence* - it is possible to argue that the failure of epistemological reconstruction and explanation is in principle a sign of ontological emergence. In other words, we are sometimes lacking the ability to reduce certain phenomena to the underlying level only because they “really” are novel also on an ontological level. Given that reality produces ontologically emergent properties, they must be within some reach of

computer simulation – even if only negatively by failing to simulate ontological emergence directly. If we don't know how any reducible relation would have to be constructed, then we could, to a certain extent, be assured that this inability - potentially in a diminishing small number of cases - demonstrates ontological emergence.

3) *Computer simulations are fundamentally involved in ontological emergence* – if ontological emergence is a fundamental property of reality and mathematics is in principle capable representing this reality, then we can assume that computer simulation should be capable to model and emulate ontological emergence in most but no all cases where ontological indirectly emergence is present.

What seems to be the most plausible of these arguments? I previously argued that proposition 1), stating that computer simulation can emulate epistemological emergence only, is the most convincing suggestion. I will outline the main argument for this conclusion based on the following assumption: *epistemologically, the encoded mathematical relations “mimic” real causal relations but are not in any sense these causalities themselves or identical with them.* The further exposition of this argument will also be an evaluation of the other propositions.

5 Ontological-Material and Mathematical-Formal Causality

The argument for epistemological emergence and computer simulation is that ontological emergence has fundamentally to do with “real” causality and not abstract mathematical causality. To make the point clearer, I will relate the argument to some definitions of causality that go back to Aristotle, namely the distinction between *material causes* and *formal causes*. *Material causes* are those from which a thing comes into existence as from its parts, constituents, substratum or materials. This reduces the explanation of causes to the parts (factors, elements, constituents, ingredients) forming the whole (system, structure, compound, complex, composite, or combination) (the part-whole causation). For example, the material cause of a table is the wood it is made of.

On the other hand, *formal causes* tell us what a thing is, that any thing is determined by the definition, form, pattern, essence, whole, synthesis, or archetype. It embraces the account of causes in terms of fundamental principles or general laws, as the whole (macrostructure) is the cause of its parts (the whole-part causation). For example, the formal cause of a statue is the idea of the sculptor, the material cause the marble it is made of.

Although there is debate on these distinctions in general, they exemplify the main argument. We can assume that ontological emergence presupposes material causes, the sum of physical ingredients down to atoms and quarks and their interactions and relations, whereas computer simulation deals with formal causes only, that is with (mathematical) patterns, laws and relations.

Therefore, to generate or replicate ontological emergence we would have to replicate reality itself - and this of course would be no model at all. There might be no way to model or capture in any abstractive way ontological emergence, be it through mathematics, language or any other abstract descriptive formal system. Ontological emergence might be an undeniable fact of the dynamics and evolution of this universe, but no abstract activity can replicate these transitions to higher levels, simply

from the fact that ontological emergence is based on all the necessary facts and real causalities down to every single physical atom that enables the emergent transition.

Every abstraction from this complex whole is in great danger to possibly cut one of the important factors for emergent transition, and from chaotic systems we might conclude that even the slightest deviation from that holistic transition can damage or alter the whole transition as such. Modifying the famous word “truth is the whole” of the German philosopher Hegel, we could state “ontological emergence is the physical whole”, and this not only in a static way at some point in time, but dynamically. For example, if we theoretically assume that the emergence of life at some point in time in the history of our planet is dynamically based on the physical evolution of the universe since the big bang, then we might imagine how difficult it would be to model this – presumably - ontological emergence. There might be more to the phase transition to life than just the reaction of some biochemical ingredients at some point in time in the history of earth, without referring to some dubious designer arguments.

From these conjectures we better understand why ontological emergence is fundamentally and by definition non-reductionist, since it is based on the whole physical process and is not compatible with any abstraction whatsoever. Yet scientific reductionism necessarily entails abstraction.

6 Simulation and Mental States

If we assume that mathematics is in most cases incapable to model ontological emergence based on its property of abstraction, what about language and logic based simulation as in artificial intelligence? What about social simulation in general, simulation based on human decisions, strategies, logic, basic ingredients and capabilities of mind and consciousness? Are we able to capture ontological emergence through social, only partial mathematical modelling?

I would argue that we do not capture ontological emergence neither by mathematical modelling of natural systems nor by simulating higher level, mind-based phenomena in social simulation. What we are capable to do is modelling different levels with the appropriate formal language but not *transitions of levels*, and that is exactly what emergence is all about; the *transition* or emergence from a lower level of causal connections to a higher level.

For example, there is still no solution or simulation model in sight that could explain or model the emergence of mental states based on neuro-physical brain states. Apart from the fact that this would mean a major philosophical breakthrough in philosophy of mind as a possible solution to the mind-body problem, it is not clear - based on the idea of ontological emergence – if this is possible in principle. Modelling of transition would mean reduction in principle, since if we want to simulate the emergence of a higher level out of lower level, we need to understand the “laws” of emergence, the variables and transitions rules that generate the properties of the higher level – and this means to generate mental states from mathematical descriptions. This might be possible in principle, given that epistemological emergence prevails – but as of yet we have no idea how this could happen. But the fundamental objection remains: the emergence – or philosophically speaking the supervenience - of mental states on physical states is presumably a case of ontological emergence and presupposes the sum of all physical facts of reality.

7 Epistemological Emergence with a Hope

To summarize the different arguments of the relation of computer simulation and different concepts of emergence, it seems to be reasonable to favour the argument that computer simulation deals in most cases, but perhaps not in all cases, with epistemological emergence only.

The “in principle” exclusion argument can be stated as follows: if there exists ontological emergence at all – of which quantum entanglement is one of the most cited candidates (Silberstein, 1999, Esfeld, 2002) – then given at least some “ontological correspondence” between mathematical constructs and reality, computer simulation should be able “in principle” to capture – by chance or rationally - ontological emergence, even if there are strong arguments against it. Obviously, these would include a “weak” interpretation of the necessity of the “physical whole” for the evolution of ontological emergence.

Still, the “in principle” argument holds up a hope that with computer simulation, even simply by negatively detecting the mathematical non-reducibility of certain emergent features, we can compute a “glimpse at the borders” of true ontological emergence that might constitute the “creativity machine” of our universe.

Consequently, we might be happy simulating the possibly small range – compared to all possible emergent levels in the ongoing history of the universe – of all mathematically accessible phenomena that now exist and leave the “simulation” of not yet born levels to our descendants.

Another interesting, more far reaching conclusion is this: if ontological emergence is true of this universe – and in this case there have and will emerge more ontological levels in the future - then it is evidently false to conclude that is possible to reach any full description of reality by mathematics alone as for example TOE’s (Theories Of Everything) in modern physics suggest. We can describe reality with mathematics very successful on the apparently lowest level, the fundamental micro-level of physics, and we understand something about the mental and conscious level, but we understand these two levels only of a possibly infinite range of other ontological levels that have and will emerge.

Coming back to computer simulation, we can summarize the central question as follows:

How far can mathematics asymptotically approximate the necessary real causality that is necessary to elicit ontological emergence – given that it is not necessary to replicate reality to emulate ontological emergence?

As a possible answer to the questions I am tempted to say: We have to simulate – and see!

References

1. Achinov, V., Fuchs, C. (eds.): Causality, Emergence, Self-Organisation a publication of the international working group on Human strategies in Complexity: Philosophical Foundations for a Theory of Evolutionary Systems, NIA-Priroda, Moskau (2003)

2. Bertuglia, C.S., Vaio, F.: *Nonlinearity, Chaos & Complexity – The Dynamics of Natural and Social Systems*. Oxford University Press, Oxford (2005)
3. Esfeld, M.: *Holismus*. In *der Philosophie des Geistes und in der Philosophie der Physik*. Suhrkamp taschenbuch wissenschaft, Frankfurt am Main (2002)
4. Kauffman, S.: *Investigations*. Oxford University Press, Oxford
5. Laughlin, R.B.: *A Different Universe – Reinventing Physics from the bottom up*. Basic Books, New York (2005)
6. Scharf, A.: *Quantum measurement and the Program for Unity of Science*. *Philosophy of Science* 60 (1989)
7. Shapiro, S.: *Thinking about mathematics – the philosophy of mathematics*. Oxford University Press, New York (2000)
8. Silberstein, M., McGeever, J.: *The search for ontological emergence*. *The Philosophy Quarterly* (1999)
9. Silberstein, M.: *Converging on emergence – Consciousness, Causation and Explanation*. *Journal of Consciousness Studies* 8(9-10), 61–98 (2001)
10. Silberstein, M.: *Reduction, Emergence and Explanation* (2002)
11. Smolin, L.: *The life of the cosmos*. Oxford University Press, Oxford (1999)

Emergence as an Explanatory Principle in Artificial Societies. Reflection on the Bottom-Up Approach to Social Theory

Martin Neumann

Bayreuth University, Institute of Philosophy,
Universitaetsstr. 30, 95447 Bayreuth, Germany
martin.neumann@uni-bayreuth.de

Abstract. This article investigates the notion of emergence in Artificial Societies. Roughly, two competing approaches to the foundations of social science exist: A micro foundation of social theory on the one hand and a notion of an emergent holistic social theory on the other. This dichotomy re-appears also in Artificial Societies. It will be argued that philosophical decisions made on the methodological level of how to interpret the concept of emergence will result in different sociological theories. This will be demonstrated by re-examining considerations on emergence undertaken by Joshua Epstein, who argues for a micro foundation of social theory. These considerations are then settled in the context of the long-lasting debates about emergence in sociology and philosophy of science. Considerations from the complexity theory and Philosophy of Science will be utilised to develop a concept of emergence which leads to the notion of an autonomous social sphere. It is demonstrated by two examples that this concept can be applied to Artificial Societies.

Keywords: Epistemology of Artificial Societies, Complexity, Emergence, Foundations of Social Theory.

1 Introduction: The Puzzle of Emergence

Artificial Societies are becoming an increasingly widely used methodological tool for investigating human societies. They provide a virtual laboratory to view the growth of social macro phenomena by the use of agent-based modelling technologies. As it is the case in laboratory experiments, simulation allows to achieve control in order to be able to test hypotheses. Social macro phenomena can be grown from the ‘bottom-up’, namely by the interaction of individual agents. For example, in the classic Sugarscape model [1] agents collect and exchange resources. Thereby the allocation of resources can be observed, which is a fundamental objective of microeconomics. This quasi experimental approach of viewing “artificial societies as laboratories” [1, p. 4] seems promising for a social science which commonly lacks such lab investigations.

In particular, the notion of emergence is widely used to denote the macro patterns that are generated in the case of these virtual experiments [2, 3]. If it is possible to “discover a concise description of the global state of the system” [4, p. 148] one can

talk of emergence. For example, the ‘emergence of role differentiation’ [5] or the evolution of the social contract [6] are global states of a system that can be grown from the bottom-up by agent-based simulation models. In the course of the simulation run, agent-based models are able to produce novel structures by local interaction according to simple rules [7–9]. Moreover, these effects are often unforeseen by the model designer himself; this proves the usefulness of these virtual laboratories. Hence, emergence is at the heart of agent-based modelling technologies. It should allow us to answer questions such as how cooperative relations among unrelated individuals emerge and become stable, or how social institutions, norms, and values evolve [10]. In summary, the promise of multi-agent simulation is seen to provide a tool for studying emergent processes in societies [11].

The focus of this paper is an examination of the contribution of the methodology of Artificial Societies for the foundations of sociological theory. However, a number of attempts have been made to clarify the concept of emergence and the notion of emergence is applied in highly different scientific disciplines such as complexity theory, sociology and the philosophy of life and mind. The aim of this paper is to link some of these different streams of the discourse on emergence such that findings from one field may illuminate others. A closer link between theory and methodology will demonstrate that epistemological considerations are highly relevant to the foundations of sociological theory; the philosophical decisions made on this point are decisive for the resulting sociological theory. This thesis will then be utilised to investigate the question as to whether Artificial Societies should aim at a micro foundation of social phenomena [e.g. 12] or whether it is possible to generate a social *sui generis* by means of Artificial Societies [13, 14]. In fact, Epstein [15] and Sawyer [16] have argued that agent-based modelling techniques imply a micro foundation of social theory. At first sight this seem to be quite straightforward: namely, agent-based models operate with individual agents. However, a rearrangement of some well-known facts from various disciplines will uncover an argument for the contrary position.

This will be done in the following steps: first, the problem of emergence will be outlined by following considerations of Epstein on sociological explanations and emergence in Artificial Societies. Next, the problem will be settled in a wider framework of the problem of emergence in sociology and the philosophical roots of this concept. Then considerations on emergence based on philosophy and complexity theory will be developed which enable a conceptual clarification of the notion of emergence in Artificial Societies to be made. This will be illustrated by two examples. Finally it will be demonstrated how Artificial Societies might uncover an emergent sociality.

The puzzle of emergence in Artificial Societies can nicely be illustrated by a notion of a generative social science developed by Epstein [15]. He holds that agent-based computational models permit a distinctive approach to social science, which he calls ‘generative’. This is closely connected to the bottom-up approach of agent-based modelling techniques: it is characterised firstly by a Generative question and secondly by a Generative experiment. The question reads as follows: “how could the decentralized local interaction of heterogeneous autonomous agents generate the given regularity?” [15, p. 41].

This question is answered by the Generative Experiment. Of course, the experiment involves running an agent-based simulation model. The general structure of the experiment is as follows:

“Situate an initial population of autonomous heterogeneous agents in a relevant spatial environment; allow them to interact according to simple local rules, and thereby generate - or grow up - the macroscopic regularity from the bottom up.” [15]

This experiment is the explanation of the phenomenon. Thus, the explanation of a social macro phenomenon is provided by generating it in the computer experiment. Now, this grown up regularity is commonly denoted by the term emergence. However, the question remains unanswered as to when to talk of emergence or “whether the emergent behaviours ... are in some sense programmed in” [4, p. 149]. The first and simplest answer to this question originated from the natural language: namely, to denote macro patterns as emergent if the local rules on the micro level are not intentionally programmed to produce these patterns. This means, referring to the surprising effects that simulation models can produce, to denote surprise as an instantiation of emergence:

“A particular loose usage of emergence simply equates it with surprising or unexpected, as when researchers are unprepared for the kind of systematic behaviour that emanates from their computers.” [1]

However, this is a subjective notion. As Epstein and Axtell already noted, if one equates emergence with ‘surprising’ it has to be asked: Surprising to whom [1, 17]? Yet this cannot be a scientific concept of emergence. This vagueness surrounding the word emergence caused Joshua Epstein [15] to undertake a more comprehensive investigation of the epistemology of emergence, asking how the classical concept of emergence could be used in generative agent-based social science.

For this purpose Epstein investigates the example of a swarm of bees creating a hive. He argues that “typical of classical emergentism would be the claim: no description of the individual bee can ever explain the emergent phenomenon of the hive” [15, p. 55]. A hive could be seen as a typical example of a stable macro pattern that may even – virtually - be generated by the means of Artificial Societies – even though in this case it is not a human society under investigation. In fact, the so-called MANTA model [18] simulated an artificial ant society. Since it is possible to discover a concise description of this macro pattern this might be seen as a typical example of an emergent phenomena; e.g. the MANTA model is able to show “the emergence of a division of labour within a nest of ‘simple’ ants” [18, p. 190] and thereby reproduce the sociogenesis of an ant colony. The same holds for the hive created by the bees: it is a macro phenomenon emerging from local interactions.

However, the question arises, as to how these patterns are generated. Obviously, an agent-based model produces the results by local interactions between the agents. According to the generative social science, these rules of interaction have to be part of the description of the individuals. In the case of the bees, these rules must include that an individual bee, put together with other bees, will create a hive. Epstein then concludes that “it is precisely the adequate description of the individual bee that explains the hive” [15]. Otherwise the definition of the bee is incomplete. Hence, the purpose

of generative social science is reduction; namely, to explain a social macro phenomenon by a precise definition of the micro level. This includes that social phenomena have to be inherent in the definition of the individuals. Yet emergence disappears; while the Generativist Motto reads ‘not generated implies not explained’, the notion of emergence implies the converse relation, namely that that the emergent is more than the sum of individual properties [15]. This is denied by Epstein. For this reason Keith Sawyer [16] also assumes the position that the methodology of agent-based modelling is a reductive approach to social theory¹. Consequently, in a later paper on ‘Remarks on the Foundations of agent-based generative Social Science’ [19], the word emergence is not mentioned.

Hence, this is the puzzle: originating from the intuitive notion that agent-based modelling provides a tool for the investigation of emergent social processes, further examination of the notion of emergence resulted in a complete rejection of emergence at all: one the one hand, it is obvious that the simulation outcomes are a result of the model assumptions. Yet, on the other hand, of what interest could the simulation be, if the social properties of the model would merely be a part of the definition of the agents? Are emergent properties necessarily programmed in?

Moreover, these methodological considerations are highly relevant to social theory: obviously, ants and bees are not human beings. However, by biological analogy, Epstein draws a sociological conclusion: namely, that any explanation of social facts should be a reduction to the level of individual actors. Hence, by discussing the philosophical question of emergentism, Epstein is pleading for micro reductive social science: The individuals are the ultimate source of social phenomena. This implies that the social sphere is a mere epiphenomenon that can be eliminated from an explanation without loss of explanatory power. Sociologically, this is the theoretical position of methodological individualism. However, this is exactly the opposite position to classical sociological emergentism. In the following, therefore, a brief outline of the long-running debate on individualism and collectivism will be given.

2 The Historical Problem Setting

2.1 Emergence in Sociology

Within sociological debates the concept of emergence appears in foundational issues. Some basic components of this concept will be outlined in this section. However, it must be emphasised that shortcomings and simplifications of this long-running debate cannot be avoided in a brief summary. For a comprehensive review compare e.g. Sawyer [16].

Broadly speaking, since the very beginning of the development of sociological theories, two major approaches towards the foundations of sociology can be distinguished: methodological individualism and methodological holism. This is one of the central distinctions in the history of sociology [22, 23], sometimes called the micro-macro distinction, methodological individualism versus methodological holism, or

¹ In fact, even in the case of bees the situation is not so simple: research on swarm intelligence [20, 21] arrived at the result that “the complexity of an individual insect is still not sufficient to explain the complexity of what social insect colonies do.” [20, p. 6].

action versus structure [4, 24–27]. In fact, in recent decades there have been numerous attempts to overcome this strict opposition, for example, Giddens' [28] structuration theory or Margaret Archer's [24] morphogenetic approach. But for the sake of argument, the two extreme positions of methodological individualism and holism are under consideration.

Individualists claim that – at least in principle – every social fact has to be explained in terms of its individual actors [12, 29–32]. This is a reductionistic approach to social reality. It is motivated by philosophical considerations about ontology [24]. In particular, reductionists claim that every entity has to be capable of being perceived by sense data. It is therefore claimed that only men of flesh and blood are observable entities [33]. Seen from this point of view, every purely sociological explanation without reference to individual actors is suspected of reification.

The holistic approaches, on the other hand, refer to a social reality with properties or laws that are not reducible to laws and theories of lower level domains such as those dealt with by psychology or biology. In particular, this point of view is connected with the idea of an autonomous science of society, namely the science of sociology. This goes back to the early protagonist of the very name sociology, August Comte [34]. Emil Boutroux introduced the idea of a hierarchy of irreducible levels of analysis [35]. In contemporary theories, this idea of an independent social level of reality is particularly elaborated in the tradition of Talcott Parson's sociological systems theory [36]. Further examples can be found in Blau's structuralism [37] or Bhaskar's social realism [38]. Some theories [e.g. 39] utilise the concept of autopoiesis to denote the autonomy of social laws.

The most crucial point that has to be clarified by these theories is the question of what social reality consists of. Hence, these theories have to say something about the relationship between social structure and individual action. Nowadays, debates on this topic usually refer to Emil Durkheim as a starting point for the debate [40]. In his famous 'rules of sociological methods', Durkheim [13] refers to society as both an entity, which is "not a mere sum of individuals" and to the idea that "Social things are actualised only through man: they are a product of human activity". At first sight this seem to be contradictory: Is society more than the mere sum of individuals or is it just the product of human activity?

However, he holds the view that the composition of individual elements generates a new level of reality. According to Durkheim, this new level of reality contains collective forces 'as real as cosmic forces' [41]. In particular, his example of collective forces determining suicide rates has become very famous. It has to be emphasised that Durkheim himself never used the term emergence but denoted this by the term *sui generis*. Yet, at this point the term emergence comes into play [16, 40]: a concept of a social reality that in fact is created by individual actors but nevertheless cannot be reduced to the actor level. It is claimed that social structure is different to the mere sum of individuals.

What is *not* provided by Durkheim and subsequent theories, however, is a theory of how exactly a new level of reality emerges. For this reason, adherents of social reality assumed a defensive position [42]. This is the case because they were unable to clarify the ontological status of the emergent phenomena. One example is the position of Goldstein, which Archer [24] calls descriptive emergence:

“No sociological theory needs to make explicit reference to sociological emergence. When methodological individualists assail this or that theory as holistic, ...its defenders have always the possibility of pointing to methodological emergence ...” [43, p. 281].

In conclusion, it has to be remarked that within sociological discourse the notion of emergence is to denote exactly the opposite of a bottom-up approach: namely, emergentism is the opposite of methodological individualism, as favoured by Epstein.

However, adherents of an emergent view on society failed to clarify the ontological status of the emergent level of social reality. Emergence is not explained but merely introduced at the point where an explanation would be required. In the following, therefore, a brief outline of the epistemological roots of emergence will be given.

2.2 Epistemology of Emergence

Obviously many details of the long-running philosophical debate on emergentism will have to be left aside for the purpose of this paper. For a more comprehensive description, readers may refer to McLaughlin [44], or Stephan [45, 46]. In short, classical emergentism can be characterised as non-reductive naturalism. This consists of the following ideas:

- 1) As a form of naturalism it rejects the existence of any non-naturalistic entities. Yet, as non-reductive naturalism, this involves a hierarchy of levels of reality which can not be reduced to one another even though they are closely dependent. Examples are the relation of the mind and brain, the relation of chemical substances and living beings, or the relation between human individuals and society [47].
- 2) The central assumption of emergentism concerns the question of how these levels are related to one another. Following Bedau, [48] the core principles of emergentism can be summarised as follows:
 - a) Emergent phenomena are somehow constituted by and generated from underlying processes.
 - b) Emergent phenomena are somehow autonomous from underlying processes.

Aspect 2a) can be found in both the sociological discourse on emergence and in Artificial Societies: As Durkheim already noted, emergent social processes are constituted by individual human beings. Also the phenomena generated in Artificial Societies are generated by agents' interactions. What is contested by Epstein and the methodological individualists is the second part of this definition, namely that the social level of reality is nevertheless autonomous from its individual actions. A historical recourse therefore might help clarify what is – or can be – the meaning of Bedau's sketchy phrase 'somehow autonomous'.

The basic ideas of the notion of emergence can already be found in the mid 19th century in the work of John Stuart Mill. In Volume III of his 'A System of Logic' [49], Mill distinguished between merely resultant causal effects of a composition of causes and a type of causation where the effect of the joint action of two or more types of causes is not the sum of the effects each type of cause would have if it had acted as

the sole causal factor [44]. One example of the former is Newton's laws of motion, where two forces acting in different directions are summed up by vector addition. One example of the latter is chemical laws emerging from physical laws or biological laws emerging from chemical laws. Yet although Mill introduced the differentiation between resultant and non-resultant, i.e. emergent causal effects with respect to chemical reactions, he was a reductionist with respect to social sciences [46].

The very terminus 'emergence' was introduced by Georg Henry Lewes [50]. Finally in the 1920s emergentism was mostly influential in the work of Samuel Alexander [51], Lloyd Morgan [52], and C.D. Broad [53]. Alexander distinguished between lower levels of existence and higher levels of existence which emerge from the lower ones. He claimed that the laws governing the emergent levels can by no means be explained in terms of the lower levels, but have to be "accepted with the natural piety of the investigator" [51, p. 46]. Examples of such emergent levels of existence are the existence of life and mind. This idea was elaborated by Lloyd Morgan: he contrasted emergentism to dualistic explanations as can be found in vitalism on the one hand as well as to what he called a mechanistic, i.e. reductionist, cosmology on the other. Thus, as Broad finally puts it, emergentism is opposed to the idea that "it would be theoretically possible to deduce the characteristic behaviour of any element from an adequate knowledge of the number and arrangement of the particles in its atoms ..." [53, p. 70]. It is important to note that emergentism does not imply that no knowledge of such a whole can be gained once it has been observed, but rather stresses that "in no case could the behaviour of a whole composed of certain constituents be predicted merely from a knowledge of the properties of these constituents, taken separately ..." [53, p. 63].

Note that this is an ontological claim. In particular, chemical synthesis served as a constant example for these early emergentists. E.g. the transparency of water is seen as emergent from its molecular components. Obviously, water is constituted by the underlying molecular level. However, the properties of water, such as its transparency, is – or was – not explainable by the properties of molecules. Hence, they are viewed as 'somehow' autonomous from the underlying level. Therefore water was regarded as a different entity to its constituent molecules.

Until the discovery of quantum mechanics, no scientific laws were known that could reduce chemical reactions to underlying atomic processes [44]. This qualifies emergentism as a philosophy of science. However, with the advent of quantum mechanics, the appeal of emergentism diminished rapidly. In particular, this led members of Logical Positivism to hold a highly sceptical view of emergentism. According to Hempel and Oppenheim [54] and Nagel [55], a property is only emergent relatively to the state of a theory; as indicated by the example of chemical reactions, a phenomenon not explainable by one theory might be explainable by another. To denote a phenomenon as emergent mainly indicates a lack of knowledge:

"Emergence is not an ontological trait inherent in some phenomena; rather it is indicative of the scope of our knowledge at a given time; thus, it has no absolute, but a relative character; and what is emergent with respect to the theories available today may lose its emergent status tomorrow." [54, p. 263]

Hence, starting as an ontological claim, scientific progress caused the notion of emergence to transform into an epistemological one. However, the situation is different now to mid 20th century chemistry: due to scientific progress, the notion of emergence is appears again. This is particularly driven by computer science and the newly available computational power.

The study of nonlinear dynamical systems, complex adaptive systems and computational theory has led to the concept of emergence in self-organising systems, which emphasises the notions of radical novelty of the emergent phenomena. These are not pre-given wholes but arise on the macro level as complex systems evolving over time [7]. Obviously, the development of Artificial Societies rests on this stream of research.

Yet, viewing Artificial Societies as a tool for studying emergent processes in societies [11] leaves open the question as to whether the emphasis is on the word ‘tool’ hence, on the epistemological side, or on ‘processes in societies’, which would stress an ontological statement.

3 Concepts of Emergence

Thus, at some point a terminological confusion remains. However, this can be dissolved by inventing a terminological distinction, developed by Stephan [56]. Following Stephan, three core principles can be identified in the concept of emergence, namely²:

- Only physical explanations in the broadest sense are introduced. In particular, entities like a *res cogitans* in the Philosophy of Mind or the vitalistic principle of vigour are rejected.
- The system’s properties depend nomologically on its microstructure. Thus, if there are no differences in the constellation of the system, no differences in the systemic properties are assumed.
- There are two kinds of properties in a system: firstly, properties that can be found in the components of the system and secondly, properties that are not properties of any components of the system. These are the emergent properties of the system.

These principles comprise the minimal conditions for emergent properties: namely, the claim that systemic properties exist [comp. also 47, 57]. They can capture the systematic points of the historical debate as outlined above. However, by adding further assumptions to these core principles, this emergentism can be expanded to what Stephan denotes as either synchronic or diachronic emergentism [56]:

² Stephan denotes the basic principles of emergence as weak emergence, which can be expanded to what he calls strong emergence by the additional features of synchronic and diachronic emergence. The same distinction between weak and strong emergence is also introduced by Mark Bedau, who will be considered later. However, Bedau denotes something completely different by the same terminology. To avoid terminological confusion, the terminology of weak and strong emergence will be avoided completely in this article.

- Synchronic emergentism claims that the emergent properties are irreducible to the properties and relations of their parts and are thus not explainable at all. For example, this is the position of Samuel Alexander³. Thus, this is a notion of emergence that can be found in the old debate on emergentism about the Philosophy of life and mind.
- Diachronic emergentism, on the other hand, stresses the novelty of the emergent features. If the further assumption is added that these phenomena are not even predictable, this concept of emergence is called structural diachronic emergentism. This is not identical with the notion of irreducibility. In fact, it is the case that irreducible phenomena are also not predictable, but the contrary is not the case. A phenomenon might be not predictable but nevertheless explainable: A typical example is the mathematical chaos theory. An essential result of chaos theory is that mathematical functions with an unpredictable behaviour exist because marginally different initial values can produce completely different trajectories. Hence, this is the notion of emergence which is applied in the 'new' debates on complexity.

In fact, the concept of emergence in sociology is a concept of synchronic emergence: The problem with which sociological explanations are faced is the synchronic determination of individual actors by the emergent level of social structure. The classic example is Durkheim's claim that suicide rates are determined by social factors, i.e. the collective forces.

Yet, within Artificial Societies the situation is different: Already the notion of 'growing' Artificial Societies indicates that there is a temporal dimension inherent in the simulation process; hence, the Generative experiment to grow up the macrostructures of interest is a case of diachronic emergence: Within the simulation experiment the emergent macrostructures are novel phenomena; they appear only in the course of the simulation. For example, in the case of the hive, this is emergent because it is built by the bees. Hence, if the concept of diachronic emergence is taken into account, there is no need to reject that it is an emergent phenomenon.

³ This is similar to a concept of so-called horizontal emergence, as introduced by Sullis [58]. Namely, the simultaneous existence of the underlying and emergent level of reality. Sullis emphasises that culture is a prominent example of horizontal emergence. In fact, this is a central problem for the social sciences: social structure exists at the same time as the actor level of social reality. However, Sullis is originated in the tradition of complex dynamical systems, which typically investigates changes of variables in the course of time, in particular the emergence of novel phenomena. Therefore for Sullis, the paradigmatic case of emergence is what Stephan calls diachronic emergence, while social scientists are more concerned with synchronic- or horizontal emergence. This is a source of confusion in debates about emergence: the fact that emergence occurs in different sciences with different traditions and different problems. Hence it is not unusual for the same terminology to sometimes denote something different, as is the case with the notion of weak and strong emergence employed by Stephan and Bedau, while in this case a similar distinction is denoted by a different terminology.

However, the question remains as to whether or not the behaviour of an agent-based simulation model is predictable⁴. If it cannot be deduced without simulation, the situation is analogous to the mathematical chaos theory; one can talk of structural diachronic emergence. This can be formulated more precisely by a concept of emergence based on Complexity Theory. Darley [59] formulated a definition of emergence based on the computation times required to derive a solution. He compared the computation times for an analytical solution and for a simulation:

Let $u(n)$ be the computation time for an analytical solution. Analogously, let $s(n)$ be the computation time for the simulation. Even though Darley is not explicit about his definition of computation time, for instance computational complexity theory could be used here. Computation time can then be defined as the number of steps used by an abstract machine – such as a Turing machine – in a particular computation. This is a standard definition of complexity classes. Then, Darley defines emergence in the following manner:

$$\begin{aligned} u(n) < s(n) &\Rightarrow \text{The system is not emergent} \\ u(n) \geq s(n) &\Rightarrow \text{The system is emergent} \end{aligned}$$

Obviously, this is a gradual concept of emergence. Moreover, there is a temporal dimension inherent in this distinction: it might be possible to find a solution for an equation which has not yet been solved⁵. However, the final limits of analytical solutions are reached if it holds that:

$$u(n) \rightarrow \infty$$

In 1936, Turing proved that such a limit of knowledge exists. His famous halting problem states that it is not decidable if a computer program will eventually halt or not. This means that it is not decidable if an analytical solution is possible at all, i.e. if it holds that $u(n) < \infty$ [60]. Hence, complexity considerations can shed light on Beaudau's sketchy phrase of 'somehow' autonomous.

This terminology can be utilised to investigate Epstein's consideration of emergence. At first sight, his stream of argumentation seems to be straightforward: Classical emergentism claims that emergent phenomena are not explainable, but within Artificial Societies emergence is the principle of an explanation. Hence, it is incompatible with classical emergentism. Nevertheless, an objection to this conclusion can already be formulated on the methodological level: If Darley's definition of emergence is applied to the above example of the definition of a bee, it is obvious that the

⁴ Of course, it is possible to predict a phenomenon once an instance of it has been observed. For example, if it is known that bees will create a hive, it is possible to predict that this event will occur in the future. If one relies on the assumption of a stable course of the world, no insight into the mechanisms that generate the phenomenon is needed. However, this is simply a rule of thumb. Some classical theories of social emergence provide an example of the description of such regularities. However, this is not a generative explanation. The question under investigation here is whether emergence is consistent with a generative bottom-up approach to social theory.

⁵ In fact, there is still even a subjective element inherent in this concept of emergence: Darley [59] cites Richard Feynman, who found a problem overwhelmingly complex, but when he explained the problem to Fermi, he was easily able to predict the solution.

hive is not an emergent phenomenon at all, given Epstein's definition of a bee: If a bee is simply defined as an entity that, put together with other bees, creates a hive, then it holds that:

$$u(n) < s(n).$$

Namely, a brief look at the definition of a bee will suffice to state that it will create a hive. A simulation of this process would then be superfluous and, in conclusion, the concept of emergence is not needed at all. However, the behaviour of the simulation models is not usually so obvious. In such cases a concept of emergence could be useful. In fact, there is already a definition of emergence for the science of complexity with simulation at the core of its argument. This is given by Mark Bedau [48]:

Macrostate P of S with microdynamic D is emergent if P can be derived from D and S's external conditions but only by simulation.

The most important aspect of this definition is the word *only*. Note, that by simulation Bedau means computer simulation. A more formal notation of this definition would thus be to rely on computational times, i.e. the case in which the computational time for an analytic solution would reach infinity:

$$u(n) \rightarrow \infty$$

Presumably, in the case of problems studied by Artificial Societies, it typically holds that $u(n) \geq s(n)$. For example, an economic equilibrium might be unstable [61] or may only be reached after a long process of convergence [62]. In these cases, it appears to be reasonable to assume that insights can only be gained by simulation and it can reasonably be neglected that the emergent results are already 'programmed in'. Moreover, the final limit is reached when the equation is analytically unsolvable. Therefore, the question remains as to whether an example of a social phenomenon can be found which fulfils this definition⁷.

⁶ He calls this weak emergence. However, this notation is different from the concept of weak emergence as introduced by Stephan.

⁷ Moreover, at this point the tricky question of downward causation comes into play: While the so-called upward causation is a causal link from the underlying level to the emergent level of reality, in the case of downward causation the causal chain works in the opposite direction. Sometimes it is claimed that emergent phenomena have a causal influence on the underlying level. A typical example is a thunderstorm: a thunderstorm is an emergent phenomenon which is generated by particles of water and air. However, the motion of these particles is determined by the emergent phenomenon of the thunderstorm. While upward causation is widely accepted as unproblematic, the notion of downward causation is regarded as problematic: Originally introduced by Campbell [63], the notion of downward causation has stimulated controversial discussions. For critical comments see, for example, Kim [64]. Other authors employ the notion of downward causation as a useful tool, in particular, for the analysis of biological phenomena. See Emmeche et al. [65]. In the case of social sciences, downward causation is introduced, for example, by Hodgson [66]. To concentrate on only one problem, namely the question of emergence, the problem of downward causation is not taken into account in this article. Thus, it is solely investigated whether it is possible to introduce the notion of an autonomous social sphere, but not if and how this social sphere is able to influence the actor level of social reality. The argument can be developed without reference to downward causation. However, it has to be emphasised that Durkheim, in his famous investigation on suicide [41], claims the existence of downward causation: Actions of individuals, namely to commit suicide, are determined by emergent collective forces.

4 Emergence in Artificial Societies

4.1 Examples

Now let us take some examples from the social sciences. The problem turns out to be the question of whether an example can be found where it can be proven that computational times for an analytic solution reach infinity. It will be demonstrated that this is crucial for the sociological theory. Namely, whether the claim that Artificial Societies provide a tool for studying emergent processes in societies is an epistemological or an ontological statement.

a) Economic equilibrium prices

In fact, Walrasian models of general equilibrium are a well-known example. They can serve as an example because on the one hand microeconomics provides one of the strongest forms of methodological individualism and on the other hand the terminology of the economic theory is highly formalised. This allows for an analytical treatment of the problem. Thus, they serve as a proof of existence. Moreover, there are simulation models which are able to generate local prices by the interaction of individual agents. The most famous model is the classical Sugarscape model. Formally, Walrasian models of general equilibrium are the mapping of a space of choice functions of consumers and producers into the space of real numbers; Thus, they are a structure of the following form [67]:

$$A = \langle \mathbb{R}^{(m+n)l}, I, J, \{X_i, x_i\}_{i \in I}, \{Y_j, h_j\}_{j \in J} \rangle$$

with the following definitions:

l : dimension of commodity space

I : cardinality m of consumers

J : cardinality n of producers

$(X_i, x_i), (Y_j, h_j)$ feasible space of alternatives.

It has been proven by Kramer [68] and Lewis [67, 69] that this problem is not solvable; i.e. $u(n) \rightarrow \infty$. For the argument it is important to note that microeconomics is a social theory purely on the actor level. In the terminology of classical emergentism provided e.g. by Broad, the objective of microeconomics is to derive the properties of a social fact $R(A,B,C)$, i.e. a ‘whole’, from the complete knowledge of the properties of the individual elements A , B , and C . Yet, the equilibrium price is a macrostructure that cannot be derived by analytical means from the actor level. Hence, it can be proven that the invisible hand is in fact invisible. Even complete knowledge of the actor level is not sufficient to predict the resulting equilibrium price. This shows the limits of a purely microsocial analysis and, in this sense, the autonomy of the structural level of social reality; i.e. the macrolevel. The notion that social phenomena should be explained ‘in principle’ in terms of individual actors [e.g. 29] can be reversed: it can be

proven that social macro phenomena exist which ‘in principle’ cannot be predicted by the means of individual actors⁸.

If the behaviour of the system cannot be predicted by the underlying assumptions then it is reasonable to apply the notion of structural diachronic emergence to it. Moreover, the fact that simulation models exist which generate unpredictable local prices demonstrate, that this example fulfils Bedau’s definition of emergence: following Bedau’s definition, prices can be denoted as the macrostate P of the market, which is the system S . The micro dynamic D of the system S can be identified by the local interactions of the actors in the market. The simulation models prove, that (semi-stable) local prices can be generated by the agent’s interactions D and S ’s external conditions by simulation. However, the proof that Walrasian equations are not analytically solvable demonstrates, that this macro state P can *only* be derived by simulation. Hence, Bedau’s definition is fulfilled: the behaviour of the system cannot be predicted, only generated.

b) Cross-validation

From the standpoint of Artificial Societies, a perhaps even more interesting example is described by Moss and Edmonds [70]: They develop the notion of cross-validation of an agent-based simulation model. In particular the notion of cross-validations implies that agent-based models, validated on the micro level, allow the generation of statistical patterns on the macro level which are in accordance with empirical observations: there are a widespread number of cases where the aggregate data exhibits unpredictable clusters of volatility. In fact, this feature can be reproduced by the means of agent based simulation models. Moss and Edmonds stress that “this result does not occur because we tune our models to produce these kind of time series but rather seems to be a consequence of the characteristics we put into our models” [70, pp. 1121 f.]. They presume that the mechanisms to arrive at this result can be identified with the social embeddedness of the actors in the agent-based model. The particular feature of agent-based models that the agents are able to interact and influence one another is responsible for this result. Note that this is a crucial difference to the actor model of the neo-classical economic theory. However, this is the mechanism which leads to increasing complexity and, in turn, to the unpredictability of the results. Thus, the model exhibits typical features of complex adaptive systems, which give rise to emergent phenomena.

Hence, the notion of cross-validation implies the assumption that validation on the micro- and macro level are independent of one another. This, however, means that the statistical macro patterns cannot be derived analytically from the design of the agents on the micro level as is indicated by Epstein’s claim about the definition of bees. If these patterns were to be a logical consequence of a micro-validated model they

⁸ In fact, Sawyer [71] has investigated a variety of mechanisms of emergence, in particular, supervenience and multiple realisability. Sawyer argues that in the case of social sciences these conditions are in fact met. In the case of Artificial Societies another mechanism is of greatest importance: the interaction of the agents. This will be demonstrated in the next example. Yet, the focus of this article is on definitions, not on mechanisms of emergence. In fact, there are also interactions which do not lead to the emergence of any phenomena. However, of importance to the argument is the fact that emergence does not necessarily imply irreducibility.

would not be unpredictable and the notion of cross-validation would make no sense. Yet such an impossibility result is not available and is presumably hard to derive at all. Nevertheless, it will presumably also not be possible to prove the contrary; the state of the art makes it highly plausible that it holds that $u(n) \geq s(n)$, i.e. that the statistical macro patterns can at best be generated by simulation. Hence, it fulfils Be-dau's definition of emergence: the statistical macro level is computationally autonomous from the micro level, since it can be derived only by simulation.

On the one hand, the things these two examples have in common is that they demonstrate the usefulness of Artificial Societies as a tool for studying social processes which exceed the ability of mathematical theory to produce analytical solutions. Hence, reference to emergence can be regarded as an epistemological statement. Yet, the sociological theory following Durkheim made an ontological claim - even though it was left unexplained.

On the other hand, they also represent features of social reality. Prices are social facts and the notion of cross-validation implies that unpredictable cluster of volatility can in fact be observed on a statistical macro level. Thus, the question still remains unanswered as to whether the notion that Artificial Societies are a 'tool for studying emergent processes in societies' is an epistemological or an ontological statement. This question will be examined in the following section.

4.2 Emergence in Artificial Societies

For an analysis of this problem, it should be taken into account that Artificial Societies provide a virtual laboratory: they enable the researcher to investigate social phenomena that otherwise might not be possible to generate. This, however, is a quite typical feature of laboratories. Hence, it might be useful to take findings from the philosophy of experimental science into account. In fact, the lab has gained a lot of attention in the past decades in the philosophy of science. Even though a by no means comprehensive review of the philosophical findings can be undertaken here [comp. e.g. 72–75], reference will be made to some arguments developed by Ian Hacking [76, 77] concerning scientific realism: in particular, he argues for laboratory realism.

Hacking explains this position by the example of the history of the atom: in the 19th century it was highly contested that electrons really exist. Phenomenological accounts described atoms as merely a useful terminus but neglected that they 'really' exist. It was admitted that the phenomena of heat and electricity exist and that the theory of atoms might help to predict some phenomena of interest. Nevertheless, atoms themselves were regarded as fictitious. Hence, they were merely regarded as an epistemological tool. However, gradually it became possible to undertake more and more experiments. In these experiments it became possible to manipulate and use atoms in a controlled manner. Atoms therefore became what Hacking calls experimental entities: they can be used by the experimentalist. Hacking regards this ability to intervene as a practical argument for realism. Thus, atoms became a part of the furniture of the world.

This leads to the sociological question: in fact, Artificial Societies provide a laboratory for sociologists. As outlined in the section on emergence in sociology, the assumption of a social reality is comparable to the situation of the ontological status of atoms in the 19th century: so far it has gained a more or less hypothetical status: by enhancing the explanation of variance the introduction of social facts is useful for

empirical research [24], but the question of what emergence consists of remained unanswered. Statistical research does not necessarily make ontological statements. However, this can be done by means of Artificial Societies: Note that theories of emergence do not claim that the emergent phenomenon is an entity separate from the processes from which it emerges. All theories of emergence start from the idea that the emergent phenomena are constituted by underlying processes [48]. In fact, Sawyer [71] demonstrates possible mechanisms. But theories of emergence claim that, nevertheless, the emergent level is autonomous from the underlying processes. This is contested by methodological individualists and the bottom-up approach to social theory.

However, by the notion of diachronic emergence, prediction and explanation have to be separated. It implies that the emergent phenomena are not predictable even though they are explainable. Again, this is well in line with emergentism: Already C.D. Broad [53] claimed, that even though complete knowledge of the underlying level might be insufficient to predict emergent phenomena before, they might be explained once their factual appearance has been observed. Note, however, that prediction is the central concept of positivistic theories of explanation [comp. 78]. This sheds light on the explanatory value of Artificial Societies: The reflexive nature of the relation between social reality and social theory and the high degrees of freedom of social systems might be possible borderlines for an account to subsume social sciences under the (classical) explanatory mode of the natural sciences, i.e. the notion of prediction as an explanation of a phenomenon. Moss and Edmonds [70] claim that the paradigm of prediction has failed completely even in the most rigorous social science, namely economics. However, this does not imply that it is impossible to gain any knowledge of emergent social relations.

Artificial Societies, as a case of diachronic emergence, show in what sense the emergent phenomena are constituted by the underlying process; simply since they can be generated from it. However, Artificial Societies show also in what sense it can be said that they are autonomous from the underlying process: its behaviour cannot be predicted. Hence, the result of complexity considerations shed light on the limits of explanatory reduction. Artificial Societies do not overcome the limitation of unpredictability. In particular, a mathematical analysis of agent-based models is usually impossible. Even if it is possible to grow phenomena on a computer screen, it is not possible to deduce the results of such a simulation experiment by purely analytical means. This is demonstrated by the examples above. By generating a phenomenon, one does not even need to fully understand the mechanisms of this process. Hence, emergence is in fact an explanatory principle; emergence is of epistemological value.

However, epistemology and ontology are closely connected: The laboratory is an epistemological tool. It is a common assumption among laboratory scientists, however, that the objects studied in the lab belong to the domain of the world. Obviously, the virtual lab on the computer screen is different from the physical lab.⁹ Yet it can provide insights into the question of how society became an autonomous level of reality:

⁹ In particular, the credibility of the assumption of the existence of an unobservable entity is greatly enhanced if it can be found in different independent laboratory experiments. For instance the Avogadro's number N can independently derived by 13 different experiments. It is a common assumption to regard this as an argument for the reality of molecular particles [79]. So far, something comparable is missing in the lab of Artificial Societies.

To recall the examples, the notion of cross-validation indicates, that the statistical features generated by local interactions of individual agents cannot be deduced without the means of generating them in a simulation experiment. They cannot even be deduced with a complete knowledge of the individual agents, i.e. the underlying level of reality. Nevertheless, the simulation results are in accordance with empirical observations. In fact, the statistical patterns are a feature of reality. Thus, Artificial Societies provide insights into phenomena that would otherwise not be accessible. Thereby they demonstrate the possibility of an emergent, unpredictable social level of reality. Hence, Artificial Societies might help to open up the black box of emergence [7]. This is the problem left unanswered by social theorists so far. Due to the complexity of its generating mechanism, social structure is irreducible to the actor level of reality. By generating macro patterns Artificial Societies allow the experimentalist to intervene in the experimental setting. Following Hacking, it can be argued that Society became an experimental entity by means of Artificial Societies.

Thus, Artificial Societies do in fact provide insights into the micro-macro dichotomy: By growing emergent phenomena, the bottom-up approach of Artificial Societies fills the explanatory gap left open in classic sociological accounts. Artificial Societies show how it is possible to clarify the notion of an autonomous social sphere¹⁰. Note, that the situation is the reverse of the situation in chemistry after the discovery of quantum mechanics. In the mid 20th century, scientific progress diminished the attractiveness of an ontological concept of emergence. In contrast, it takes scientific progress to show e.g. the unsolvability of the microeconomic equation. The concept of emergence is of growing interest due to scientific progress.

5 Conclusion and Perspectives

Obviously, to rely on structural diachronic emergence is a considerably weak notion of autonomy. It is not claimed that social structure cannot be explained. In fact, this can be done by means of Artificial Societies. It is merely claimed that the shape of the emerging social structure cannot be predicted by means of individual actors. This holds even for the case of complete knowledge of individual agents. This is indicated by the term ‘structural diachronic emergence’. Yet prediction and explanation have to be differentiated. While it is possible to explain an emergent feature *ex post*, it is impossible to predict it *ex ante*. Thereby the emergent feature gains its autonomy:

¹⁰ The result of these considerations is nearly the opposite of the conclusion drawn by Epstein; However, it shall be emphasized that these findings might not be as contrary to Epstein’s account as this theoretical conclusion seems to indicate: Epstein defines what he calls a ‘hard social problem’ [15, p. 50] exactly by considerations about computational time and claims that “there are social problems that are undecidable in principle” [15, p. 50]. However, given his definition of a bee as a paradigmatic example of a sociological explanation, it could be contested that hard social problems exist at all. However, the example of the computation of an equilibrium price demonstrates their existence. Nevertheless, by separating prediction and explanation, the sociological conclusion is the opposite: a phenomenon like an equilibrium price cannot be deduced by analytical means from terms of the level of individual actors. It is actually an autonomous emergent phenomenon. Agent-based simulations models are a means to study structural diachronic emergence.

Since it is impossible to predict it, it is impossible to deduce it from underlying processes. In contrast to reductive explanations in chemistry in the mid 20th century, it takes scientific progress to prove this impossibility result. This concept of emergence is different to the defensive usage of emergence in classical sociological theories, which Archer denotes as descriptive emergence: these theories introduce the notion of emergence to explain macro structures only in terms of other macro structures. However, this difference demonstrates how different streams of thought may illuminate each other. Here, emergence is made explicit.

Finally, separating prediction and explanation sheds light on the scientific status of Artificial Societies: Within the humanities it is a common notion to differentiate between 'explaining' and 'understanding'. While science is concerned with explanation, the humanities aim to understand their object of investigation. However, the process of understanding does not yield a prediction of events, as an explanation in the domain of the sciences does. Hence, with respect to their explanatory principle, namely emergence, the investigation of Artificial Societies can be regarded as a contribution to the humanities.

References

1. Epstein, J., Axtell, R.: *Growing Artificial Societies. Social Science from the Bottom Up*. MIT Press, Cambridge (1996)
2. Gilbert, N., Troitzsch, K.G.: *Simulation for the Social Scientist*. Open University Press, Buckingham (1999)
3. Gilbert, N.: *Varieties of Emergence*. In: *Proceedings of Agent 2002 Conference: Social Agents: ecology, exchange and evolution*, Chicago (2002)
4. Gilbert, N.: *Emergence in social simulation*. In: Gilbert, N., Conte, R. (eds.) *Artificial Societies: The computer simulation of social life*. UCL Press, London (1995)
5. Eguiluz, V., Zimmermann, M., Cela-Conde, C., Miguel, M.: *Cooperation and the Emergence of Role Differentiation in the Dynamics of Social Networks*. *American Journal of Sociology* 110(4), 977–1008 (2005)
6. Skyrms, B.: *Evolution of the social contract*. Cambridge University Press, Cambridge (1996)
7. Goldstein, J.A.: *Emergence as a construct: History and Issues*. *Emergence* 1(1), 49–72 (1999)
8. Richardson, K.: *On the Limits of bottom-up computer simulation: Towards a nonlinear modelling culture*. In: *Proceedings of the 36th Hawaiian International Conference on System Science*. IEEE, California (2003)
9. Richardson, K.: *On the Relativity of Recognising the Products of Emergence and the Nature of Physical Hierarchy*. In: *Proceedings of the 2nd Biennial International Seminar on the Philosophical, Epistemological and Methodological Implications of Complexity Theory*, Havana International Conference Center, Cuba, January 7-10 (2004)
10. Kohler, T.: *Putting Social Science together again*. In: Kohler, T., Gummerman, G. (eds.) *Dynamics in Human and Primate Societies: agent-based Modelling of Social and Spatial Processes*. Oxford University Press, Oxford (1999)
11. Drogul, A., Ferber, J.: *Multi-agent simulation as a tool for studying emergent processes in societies*. In: Gilbert, N., Doran, J. (eds.) *Simulating Societies. The computer simulation of social phenomena*. UCL Press, London (1994)

12. Watkins, J.W.N.: Historical Explanations in the Social Sciences. *The British Journal for the Philosophy of Science* 8(30), 104–117 (1957)
13. Durkheim, E.: *Les Regles de la Methode Sociologique*. Presses Universitaires de France, Paris (1895)
14. Durkheim, E.: Representations individuelles et Representations collectives. In: Durkheim, E. (ed.) *Sociologie et Philosophie*. Libraire Felix Alcan, Paris (1924[1898])
15. Epstein, J.: Agent-Based Computational Models and Generative Social Science. *Complexity* 4(5), 41–60 (1999)
16. Sawyer, K.: *Social Emergence: Societies as Complex Systems*. Cambridge University Press, Cambridge (2005)
17. Axelrod, R.: Advancing the Art of Simulation in the Social Sciences. In: Conte, R., Hegselmann, R., Terna, P. (eds.) *Simulating Social Phenomena*. Springer, Berlin (1997)
18. Drogul, A., Corbara, B., Lalande, S.: MANTA: new experimental results on the emergence of (artificial) ant societies. In: Gilbert, N., Conte, R. (eds.) *Artificial Societies: the computer simulation of social life*. UCL Press, London (1995)
19. Epstein, J.: Remarks on the Foundations of agent-based Generative Social Science. CSED Working Paper No. 41 (2005)
20. Bonabeau, E., Dorigio, M., Theraulaz, G.: *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, Oxford (1999)
21. Kennedy, J., Eberhart, R.: *Swarm Intelligence*. Academic Press, San Diego (2001)
22. Knorr-Cetina, K., Cicourel, A.V. (eds.): *Advances in social theory and methodology*. Routledge, London (1981)
23. Alexander, J.: Action and its environment. In: Alexander, J., Giesen, B., Smelser, N. (eds.) *The Micro-Macro Link*. University of California Press, Berkley (1987)
24. Archer, M.: *Realist social theory: the morphogenetic approach*. Cambridge University Press, Cambridge (1995)
25. Mayntz, R.: Individuelles Handeln und gesellschaftliche Ereignisse: Zur Mikro-Makro-Problematik in den Sozialwissenschaften. MPIfG Working Paper 99/5 (1999)
26. Sawyer, K.: Artificial Societies: Multiagent Systems and the Micro-Macro Link in Sociological Theory. *Sociological Methods and Research* 31(3), 325–363 (2003)
27. Heintz, B.: Emergenz und Reduktion. *Neue Perspektiven auf das Mikro-Makro-Problem*. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 56(1), 1–31 (2004)
28. Giddens, A.: *The constitution of Society: Outline of the Theory of Structuration*. Polity Press, Cambridge (1984)
29. Collins, R.: On the Microfoundations of Macrosociology. *American Journal of Sociology* 86(5), 984–1014 (1981)
30. Elster, J.: *Rational Choice*. Blackwell, Oxford (1986)
31. Coleman, J.: *Foundations of social theory*, vol. 1. Belknap Press, Cambridge (1990)
32. Esser, H.: *Soziologie: allgemeine Grundlagen*. Frankfurt a.M., Campus (1993)
33. Homans, G.: Bringing man back in. *American Sociological Review* 29(5), 809–818 (1964)
34. Comte, A.: *Système de Politique Positive*, vol. 2. Otto Zeller, Osnabrueck (1967[1851])
35. Boutroux, E.: *The Contingency of the Laws of Nature*. Open Court Publishing Company, Chicago (1916[1874])
36. Parsons, T.: *The structure of social action*. The free Press, Glencoe (1937)
37. Blau, P.: A Macrosociological Theory of Social Structure. *American Journal of Sociology* 83(1), 26–54 (1977)
38. Bhaskar, R.: *The Possibility of Naturalism*. Routledge, New York (1975)
39. Luhmann, N.: *Soziale Systeme. Grundriß einer allgemeinen Theorie*. Surkamp, Frankfurt a. M. (1984)

40. Sawyer, K.: Durkheim's Dilemma: Towards a Sociology of Emergence. *Sociological Theory* 20(2), 227–247 (2002)
41. Durkheim, E.: *Le Suicide*. Libraire Felix Alcan, Paris (1930[1897])
42. Gellner, E.: Holism versus individualism. In: Brodbeck, M. (ed.) *Readings in the Philosophy of the Social Sciences*. MacMillan, New York (1971)
43. Goldstein, L.J.: Two theses of methodological individualism. In: O'Neill, J. (ed.) *Modes of Individualism and Collectivism*. Heinemann, London (1973)
44. McLaughlin, B.: The Rise and Fall of British Emergentism. In: Beckermann, A., Flor, H., Kim, J. (eds.) *Emergence or Reduction?* De Gruyter, Berlin (1992)
45. Stephan, A.: Emergence – A Systematic view on its historical facets. In: Beckermann, A., Flohr, H., Kim, J. (eds.) *Emergence or reduction? Essays on the Prospects of Nonreductive Naturalism*. De Gruyter, New York (1992)
46. Stephan, A.: *Emergenz: von der Unvorhersagbarkeit zur Selbstorganisation*. Dresden Univ. Press, Dresden (1999)
47. Bunge, M.: Emergence and the mind. *Neuroscience* 2(4), 501–509 (1977)
48. Bedau, M.: Weak Emergence. In: Tomberlin, J. (ed.) *Philosophical Perspectives: Mind, Causation, and World*, vol. 11. Blackwell, Malden (1997)
49. Mill, J.S.: *A System of Logic: ratiocinative and inductive*. Hawaii Univ. Press of the Pacific, Honolulu (2002 [1891]) (reprint of the 1891 ed.)
50. Lewes, G.: *Problems of Life and Mind*, vol. 2. Kegan Paul, London (1875)
51. Alexander, S.: *Space, Time, and Deity*, vol. II. MacMillan, London (1920)
52. Morgan, L.: *Emergent Evolution*. William & Norgate, London (1923)
53. Broad, C.D.: *The Mind and its Place in Nature*. Routledge, London (1925)
54. Hempel, C., Oppenheim, P.: Studies in the Logic of Explanation. *Philosophy of Science* 15(2), 135–175 (1948)
55. Nagel, E.: *The Structure of Science*. Harcourt, New York (1961)
56. Stephan, A.: Varieties of Emergentism. *Evolution and Cognition* 5(1), 49–59 (1999)
57. Bunge, M.: Treatise on Basic Philosophy. *Ontology II*, vol. 4. Reidel, Dordrecht (1977)
58. Sullis, W.: Archetypical dynamical systems and semantic frames in vertical and horizontal emergence. *Emergence: Complexity and Organisation* 6(3), 52–64 (2004)
59. Darley, V.: Emergent Phenomena and Complexity. In: Brooks, R., Maes, P. (eds.) *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*. MIT Press, Cambridge (1994)
60. Chaitin, G.: A Century of Controversy over the Foundations of Mathematics. *Complexity* 5(5), 12–21 (2000)
61. Canning, D.: *Evolution of Cooperation through Inter-Group Conflict*. Working Paper, Department of Economics, Queens University of Belfast, Belfast (1995)
62. Young, H.P.: The Evolution of Conventions. *Econometrica* 61(1), 57–84 (1993)
63. Campbell, D.: Downward Causation in Hierarchically Organised Biological Systems. In: Ayala, F., Dobzhansky, T. (eds.) *Studies in the Philosophy of Biology*. MacMillan, London (1974)
64. Kim, J.: Downward Causation in Emergentism and Nonreductive Physicalism. In: Beckermann, A., Flor, H., Kim, J. (eds.) *Emergence or Reduction?* De Gruyter, Berlin (1992)
65. Emmeche, C., Køppe, S., Stjernfelt, F.: Levels, Emergence, and three types of downward causation. In: Bøgh, A.P., Emmeche, C., Finnemann, N.O., Voetmann Christiansen, P. (eds.) *Downward Causation. Minds, Bodies and Matter*. Århus University Press, Århus (2000)

66. Hodgson, G.: Reconstitutive downward causation: Social Structure and the development of human agency. In: Fullbrook, E. (ed.) *Intersubjectivity in Economics*. Routledge, London (2002)
67. Lewis, A.A.: On Turing Degrees of Walrasian Models and a General Impossibility Result in the Theory of Decision Making. *Mathematical Social Science* 24(3), 141–171 (1992)
68. Kramer, G.: An impossibility result concerning the theory of decision making. Cowles Foundation Reprint, No. 274, Yale University (1974)
69. Lewis, A.A.: On Effectively Computable Realizations of Choice Functions. *Mathematical Social Science* 10(1), 43–80 (1985)
70. Moss, S., Edmonds, B.: Sociology and Simulation: Statistical and Qualitative Cross-Validation. *American Journal of Sociology* 110(4), 1095–1131 (2005)
71. Sawyer, K.: The Mechanisms of Emergence. *Philosophy of the Social Sciences* 34(2), 260–282 (2003)
72. Latour, B., Woolgar, S.: *Laboratory Life*, vol. *The Social Construction of Scientific Facts*. Sage Publications, Beverly Hills (1979)
73. Galison, P.: *How experiments End*. University of Chicago Press, Chicago (1987)
74. Giere, R.: *Explaining Science. A Cognitive Approach*. University of Chicago Press, Chicago (1988)
75. Kitcher, P.: *The Advancement of Science. Science without Legend, Objectivity without Illusions*. Oxford University Press, Oxford (1993)
76. Hacking, I.: *Representing and intervening. Introductory Topics in the Philosophy of Natural Science*. Cambridge Univ. Press, Cambridge (1983)
77. Hacking, I.: The participant irrealist at large in the Laboratory. *British Journal of the Philosophy of Science* 39(3), 277–294 (1988)
78. Stone, M.: Chaos, Prediction and Laplacian determinism. *American Philosophical Quarterly* 26(2), 123–131 (1989)
79. Salmon, W.: Four decades of scientific explanation. In: Kitcher, P., Salmon, W. (eds.) *Scientific Explanation. Minnesota Studies of the Philosophy of Science*, vol. 8. University of Minnesota Press, Minneapolis (1989)

Reconstruction Failures: Questioning Level Design

Camille Roth

Centre d'Analyse et de Mathématiques Sociales
CNRS/EHESS
54, bd Raspail, 75006 Paris, France
and
CREA
Ecole Polytechnique/CNRS
1, Rue Descartes, 75005 Paris, France
roth@ehess.fr

Abstract. In front of unsuccessful models and simulations, we suggest that reductionist and emergentist attitudes may make it harder to detect ill-conceived modeling ontology and subsequent epistemological dead-ends. We argue that some high-level phenomena just cannot be explained and reconstructed from insufficiently informative lower levels. This eventually requires a fundamental viewpoint change in not only low-level dynamics but also in the design of low-level objects themselves, considering distinct levels of description as just distinct observations on a single process.

Keywords: Modeling Methodology, Reconstruction, Emergence, Downward Causation, Complex Systems.

1 Introduction

Models aim at rebuilding certain aspects of real phenomena using some given level of description. Complex systems science in particular endeavors at rebuilding complex high-level behavior from more simple, more reliable and better-understood “atomic” mechanisms at an allegedly lower level. More precisely, with the help of low-level descriptions, it aims at (i) checking whether some already-known high-level descriptions are properly reconstructed (validation of higher-level phenomena), or (ii) discovering new high-level descriptions (new unexpected and potentially counterintuitive phenomena). This attitude has a crucial epistemological advantage over strictly high-level descriptions: it works with *simpler* and, often, *more reliable* mechanisms, thus complying with Occam’s razor law. Simulation-based models are frequently used, as analytical solutions are seldom available and limited to singular, possibly non-realistic hypotheses. In turn, the simulated system should correctly render the evolution of a selection of high-level stylized facts. To this end, a reductionist attitude is usually adopted; in other words, modeling efforts are focused on low-level items only — for instance, when attempting to rebuild psychological laws by iterating neural activity, the simulation relies on neuron-based properties and dynamics in order to reproduce psychological properties and dynamics, which are then traditionally said to “emerge”.

Here, we intend to review and comment the appraisal of unsuccessful models and corresponding simulations, then discuss in a broader framework the epistemological consequences of failed reconstructions on model and level design.

2 Reductionist Approach

2.1 Micro-founding the Higher Levels

In a reductionist setting, models rely on low-level items which are thus in charge of the whole reconstruction. This approach discards theories of the higher level to focus on “micro-founded” science — as such, it discards all impermeability between scientific fields. For instance, instead of using laws and theories of psychology, one may be willing to rebuild them by iterating the activity of neurons, which compose here the lower level, governed by biological laws (this is a current issue in computational neuroscience, e.g. for explaining adaptive change capabilities from neural plasticity, see Destexhe & Marder, 2004).

High-level properties must therefore be first translated into low-level properties by a mapping function P expressing the higher level H from the lower level L , such that $P(L)=H$. Without P , it would be unlikely to expect saying anything about high-level phenomena by just playing with low-level items. The higher level H may in turn be easily or directly described in the “natural language” of a given discipline (psychological mechanisms, sociological features such as communities), whereas L usually corresponds to more formal and simple descriptions (neural states, relationships between agents).

To achieve successful reconstruction, low-level dynamics observed through P must be consistent with higher-level dynamics, that is, a sequence of low-level states mapped by P should correspond to a valid sequence of high-level states. More formally,¹ denoting by λ (resp. η) the dynamics or transfer function of a low-level state L (resp. high-level state H) to another one L' (resp. H') — in short, $\lambda(L)=L'$, $\eta(H)=H'$ — this means that P must form a *commutative* diagram with λ and η such that (Rueger, 2000, Nilsson-Jacobi, 2005, Turner & Stepney, 2005):

$$P \circ \lambda = \eta \circ P \tag{1}$$

Indeed, the left side of Eq. 1 is the high-level result of a low-level dynamics, while the right side yields the direct outcome of a high-level dynamics. The aim of the reconstruction is to equate the latter with the former (see Fig. 1–left).

2.2 Commutative Reconstruction

Commutativity is the cornerstone of the process: should this property not hold, the reconstruction would fail. How to check it? Since P is a definition stemming from the ontological design of levels L & H , and λ is designed by the modeler, η is truly the benchmark of the reconstruction. There are nevertheless two ways of considering η :

¹ Although formulated in a specific way, this formalism could be easily transposed for a wide range of kinds of dynamics, discrete or continuous.

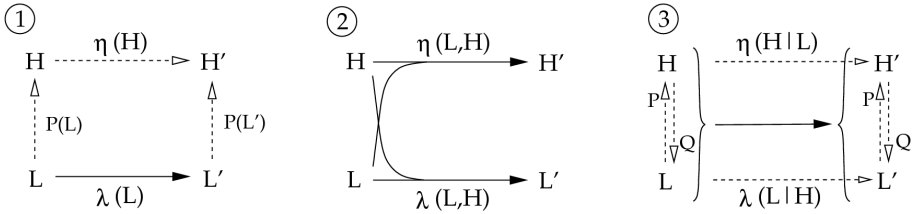


Fig. 1. Relationships between levels and their dynamics in the case of (1) reductionism, (2) emergentism (or dualism), and (3) “observationism”, or levels as observations (solid lines: causal links, dashed lines: informational dependence)

- (i) either η stems from *a priori* knowledge of higher-level dynamics – it comes from a series of measurements or a well-established theory (e.g., “can we rebuild these Zipf laws arising in that context? ”); or
- (ii) η is discovered *a posteriori* from the model – it comes from the observation of the model (e.g. “what unexpected phenomena may emerge? are they empirically valid? ”).

Verifying Eq. 1 in the first case simply refers to a successful reduction, while in the second case it induces new knowledge for the scientist, because the challenge is to exhibit a novel solution η to Eq. 1, knowing P and λ ; and then to test this stylized theoretical solution against reality.²

The success of the reconstruction endeavor depends on the capacity of “ $P \circ \lambda$ ” to rebuild η — Eq. 1 should hold in any case. This argument actually remains valid whether the underlying model is simulation-based or purely analytical: either, rarely, analytical proofs are available (e.g. gas temperature reduced to molecular interactions), or, more often, if an analytical resolution is hardly tractable, simulation methods are invoked (only proofs on statistically sufficient simulation sets are possible, using several initial states L). This is plausibly an empiricist attitude, yet each simulation is a proof on a particular case (Epstein, 2005) so the reconstruction may be

² In more details: in the first case “(i)”, consider an example where one already knows the empirical dynamics η^e of a given law of city size distribution ($\eta^e(H)=H'$, where both H and H' follow Zipf laws) (Pumain, 2004). The high-level state H is composed by P of low-level objects (cities and their populations) whose dynamics is deemed to be λ . Initially, $P(L)=H$. Suppose now that $P \circ \lambda(L)=H'$: if $H'=H'$, $P \circ \lambda = \eta^e \circ P$, the reconstruction succeeded, otherwise it failed. Validity is appraised through a comparison of $\eta(H)$ and $\eta^e(H)$?. In the second case “(ii)”, consider an example where one wants to observe the adoption rate of an innovation (a high-level dynamics) from low-level agent interactions (Deroian, 2002). Here also, P and λ are defined by the modeler, only η is induced by assuming the commutativity, i.e. this comes to find a η that satisfies Eq. 1. Often, this approach stops here: it rests on the stylized high-level dynamics η deduced from the interplay of P and λ . But at this point it should be straightforward to compare the empirically measured η^e to the modeled η – at this point, this is equivalent to the kind of empirical validations carried out in the first case: “does $\eta(H)=\eta^e(H)$?”.

considered a success as long as Eq. 1 holds true for statistically enough particular cases.³

Hence, for η is the objective of the reconstruction, when commutativity does not hold, because we assume η to be empirically appraised, the failure must be due either to λ or to P . Suppose that we stick to the fact that H is always correctly described by $P(L)$, i.e. the fact that the correspondance between low- and high-level items is valid.⁴ Then λ must be jeopardized. In this case the low-level dynamics entails, through P , a high-level dynamics different from the one given by η : the model provided by λ fails by missing something and $\lambda(L)$ is invalid, otherwise $P(L)$ would equate H .

Solutions consist in improving the description of the low-level dynamics until commutativity holds. In this paradigm, reductionism could fail only for practical reasons, for instance if commutativity requires an intractably complicated λ .⁵

3 Emergentist Approach

3.1 Alleging an Independent Higher Level

Despite this, it may also be that reductionism fails for essential reasons: even with an ideally perfect knowledge of λ , reconstruction attempts fail because H is inobservable from L : “Psychology is not applied biology” (Anderson, 1972). Here the whole is said to be more than its parts, and H enjoys some sort of independence, even when acknowledging that everything ‘concretely’ stems from the lower level (or is grounded in it).

This refers traditionally to the *emergentist* position (Humphreys, 1997, Kim, 1999).

The point is to bridge the possible failures of reductionism: the higher level is not reducible, the whole is more than the sum of its parts, even in theory; but it is grounded in the lower level so it needs to *emerge* from the lower level (Humphreys, 1997). No dualism is supposed a priori, but the cumulated, aggregated action of small objects somehow leads to the emergence of *novel* higher-level objects irreducible to lower-level objects: as Hüttemann (2005) underlines, this position “is meant to capture the intuition that there might be some sense in which the behavior of a compound system is independent vis-à-vis the behavior of the parts.” In many cases where reductionism actually fails in spite of a solid, credible and/or reliable λ , complex system methodology tends to agree with this emergentist stance.

What to do with such “irreducible” emergent phenomenon when dealing with simulation-based or analytical models? Either one considers that the emergent phenomenon

³ Equivalently, an analytical solution can be considered as simulations whose outcome is known a priori. For an extensive discussion on the wide spectrum of criteria, accurate or less accurate, that make a simulation-based model successful, see (Küppers & Lenhard, 2005).

⁴ I.e., $P(L)=H$ for all empirically valid couple of low- and high-level states (L,H). Note that this is necessarily the case when H describes higher-level patterns on L: this is what some authors seem to call second-order properties (Kim, 1998).

⁵ For the sake of practicality some models also assume that λ depends on H. However, because $P(L)=H$, which amounts to nothing more than repeat that λ depends on L, through the instrumental “simplifier” P.

has no causal power, making it a mere epiphenomenon.⁶ Yet, by being irreducible to what is being modeled at the lower-level, it is unlikely that a model, *a fortiori* a computer-run simulation, could inform us about such an epiphenomenon. One would then consider that the emergent phenomenon has causal powers onto the lower level. But again, it is unlikely that a simulation would help us in this regard: this would mean that a computer program creates something that in turn has an effect onto the program and therefore the code which the modeler has instructed the computer to execute. Likewise, a model itself would not be eager to exhibit such causal feedback – this would in fact be as if someone, writing and deriving equations, was perturbed in *the very writing of these equations* by an invisible hand adding some formulas here and there.

We would thus hope that neither a simulation nor actually any kind of model suffers “downward causation” (Campbell, 1974), which would hence be bound to exist in the real world only.⁷

Now, to take into account emergent phenomena, the modeler would have to design influences between both levels; in other terms, η would be enriched to take L into account, $\eta(L,H)=H'$, and λ would take H into account by assuming downward causation: $\lambda(L,H)=L'$ – see case (2) on Fig. 1.⁸ However then, assuming both that (i) the lower level causes high-level phenomena which (ii) have in turn a downward influence on low-level objects is nonetheless likely to raise inconsistency issues regarding low-level property violations.

Indeed, in an emergentist framework, where both upward and downward causations are present, interactions of low-level items create a higher-level object which in turn, is supposed to have an influence on the lower-level items ($L \rightarrow H \rightarrow L'$). For instance, cell interactions produce some emergent psychological feature (e.g. stress) which in turn induces biological changes (blood pressure increase). Similarly, consciousness is considered causally efficient on the activity of the body from which it is believed to originate (Thompson & Varela, 2001). Quite widely spread, this conception could be surprising: even *in vivo*, can a lower level create a higher level which in

⁶ The argument is fundamentally as follows: denoting lower-level states by “L” and higher-level states by “H”, L causes L', however at the same time L causes H and L' causes H'; so why would we need H and H' for? These two properties seem in fact merely epiphenomenal. Thus, “[i]f emergent properties exist, they are causally, and hence explanatorily, inert and therefore largely useless for the purposes of causal/explanatory theories” (Kim, 1999). But then, epiphenomenality does not differ much from reductionism, and as Bitbol (2006) emphasizes, “emergentists are inclined to require productive causal powers of the emergent properties on the basic properties.”

⁷ Donald Campbell, who introduced the term ‘downward causation’, “[a]ll processes at the lower levels of a hierarchy are restrained by and act in conformity to the laws of the higher levels” (Campbell, 1974): downward causation corresponds to the situation where a system of objects which integrates into a larger whole is in turn affected by the larger whole. More precisely, Campbell illustrates this idea as follows: “The organisational levels of *molecule, cell, tissue, organ, organism, breeding population, species*, in some instances *social system* (...) are accepted as factual realities rather than as arbitrary conveniences of classification, with each of the higher orders organising the real units of the lower level” (ibid.).

⁸ In such an emergentist setting, models then simulations would thus attempt to intertwine high-level objects in the low-level dynamics in order to rebuild emergent phenomena. This is formally close to dualism, at least in the simulation implementation.

turn influences the lower level? Detractors of emergentist downward causation argue essentially that it is redundant and, even worse, that it violates the causal rules defining the lower level; hence, they suggest, a critically erroneous principle — see e.g. (Emmeche et al., 2000).

In terms of model design, this suggests that one is likely to model something that is not even *causally* valid in the real world. As Bitbol (2006) sums up:

“Consider the crucial case of “downward causation”, namely causation from the emergent level to a basic level : from the social to the mental level; from the mental to the biological level; and from the biological to the physical level. Within their predominantly substantialist framework of thought, the emergentists are inclined to require productive causal powers of the emergent properties on the basic properties. And nothing of the sort is in sight. At most, one can find ways of seeing some complex mutual interactions of large numbers of basic components as “trans-scale” causation.”

3.2 Levels as Observations

To avoid strictly dualist models and causal & epistemological concerns, one should consider that properties *at any level* are instead the result of an observational operation (Bonabeau & Dessalles, 1997, Gershenson & Heylighen, 2003, Bitbol, 2006): the only emergence is that of several modes of access to a same process, where each observation level may provide overlapping information. The higher level may indeed yield sufficient information about an underlying process, so that we can have an idea of what happens *and* what does not happen at the lower level, and vice-versa. For example, when some individual expresses some stress (a psychological observation), one could guess that the blood pressure is higher (a biological observation). Information gained from the observation of some level *specifies* the dynamics of another level, and dynamics could be rewritten as $\lambda(L|H)=L'$ and $\eta(H|L)=H'$ — see Fig. 1. Knowing for instance that the higher level has a far slower time-scale, one could fruitfully bind the low-level dynamics to some high-level parameters, thereby significantly easing the understanding of the low-level dynamics.

Obviously, high-level reconstruction strictly from valid low-level models is possible only when the higher level is *deducible* from the lower level. When the reconstruction fails despite robust λ and η , one must envisage that the chosen lower level L does not yield *enough* information about H , and we have to check whether we are not missing something crucial when designing levels. Lane (2006) underlines this effect with a striking metaphor about “details”: there is basically no use trying to explain crises from dynamics on social classes, when the relevant item that is informative of the high-level crisis is actually at a very lower level concerning individual action. In other words, sometimes there are details that may *account for* the high-level dynamics such that the chosen decomposition into a lower-level dynamics is *essentially* inefficient for high-level prediction. Here, it may simply be that observing L will *never* yield enough information about H , and this bears identical consequences for modeling.

4 Questioning Level Design

This leads to a significant change in viewpoint: first, there is no “substantial” reality of levels, which a simulation is allegedly trying to reproduce, but an observational reality only.⁹ Consequently, there is no reciprocal causation between levels, but simply informational links: higher and lower levels are simultaneous observations of a same underlying process that *may or may not* yield overlapping information about other levels. Most importantly, some phenomena cannot be rebuilt from some given lower level descriptions – not because of higher level irreducibility but because of an essential deficiency of the lower level description.

Put differently, it is not that the whole is more than its parts, it is that the whole we are observing at a higher level is more than *these* parts we focused on.

Slightly paraphrasing the way Bedau (1997) presents the puzzle of emergence, our argument suggests that if an emergent phenomenon is somehow autonomous from underlying processes, then this emergent phenomenon is constituted and generated *not only* by these underlying processes. In this respect, reductionism makes the intuitive yet audacious bet that there is a ultimate level which yields enough information about any other “higher” level, at least in principle — which, when it works in some particular cases, gives the impression that a high-level phenomenon is *reducible*, while in fact it is simply *fully deducible*.

More to the point, what should happen when simulating, for instance, neural activity in order to provoke the emergence of a psychological phenomenon like learning, while in fact there are crucial data in glial cells which would make such attempt irremediably unsuccessful (Pfriegeer & Barres, 1996)? As such, emergentism could be a dangerous modeling approach. Yet, reductionism would not be more helpful by assuming the existence of a lowest level for which mapping functions P towards any higher level do exist. When neurons are the lowest level, attempting to model the emergence of learning could also be a problem.

Similarly and to provide another example, a social network model ignoring crucial semantic features which in fact determine real-world interactions is likely to enjoy a limited success. It is not infrequent that some social network-based community emergence model seeks to reconstruct knowledge communities without having recourse to any semantic space. In this case, “social glial cells” may just have been ignored. In contrast, what constitutes the vocabulary and the grammar of the corresponding simulations—agents, interactions, artifacts, etc.—may well need to be enriched in order to explain several key features in e.g. knowledge-based social networks. Yet, the belief that a social network is obviously sufficient to reconstruct many real-world social structures seems to be widespread, even when such attempts appear to require incredibly and possibly unrealistically complicated dynamics (Roth, 2007).

Rethinking Levels. Therefore, it may be mandatory to rethink levels. A quite frequent need is that of a third level, intermediary between higher and lower levels: a “meso-level” deemed more informative than the macro-level while more assessable than the micro-level; sometimes crucial to understand some types of phenomena

⁹ This situation is moreover clearly consistent with the means of a simulation: all significant operations are indeed happening *in silico*.

(Laughlin *et al.*, 2000). A triad of macro-, meso- and micro-levels seems rather arbitrary, and one may well imagine that some research topics involve even more levels (such as e.g. studying a (i) system of (ii) cities made of (iii) coalitions of (iv) agents who are (v) learning neural networks). In contrast, introducing new levels could also be simply more convenient — and this, harmlessly and at no epistemological cost, because levels are merely observations. Rather than claiming that each level exists as such, substantially, this approach simply claims that observation devices exist as such.

Now, how to design new levels? Various authors support the idea that introducing a new level is interesting insofar as it makes possible a better understanding and/or prediction of the system (Crutchfield, 1994, Clark, 1996, Shalizi, 2001, Gershenson & Heylighen, 2003, McGregor & Fernando, 2005). More precisely, the argument is essentially that emergent properties are high-level properties that “are ‘easier to follow,’ or ‘simplify the description,’ or otherwise make our life, as creatures attempting to understand the world around us, at least a little easier” (Shalizi, 2001). This calls clearly for choosing an observation level that provides easily key information on a given phenomenon. Here, instead of considering (emergent) high-level properties as something complicated, impossible to understand, or even irreducible – a negative and slippery definition – this informational attitude regards the high-level as something that must enable a more convenient understanding and prediction of the phenomenon — a positive definition.

This stance is very enlightening theoretically: in order to give meaning to complex systems we design new observational instruments and description grammars which help reduce reality dimensions and complexity. Going further operationally, compelling methods (Crutchfield, 1994) and effective algorithms (Shalizi & Shalizi, 2004) have been proposed to find and build automatically & endogenously a new level of observation (i) based on low-level phenomena and (ii) simplifying their description. In any case, these tools appear to be powerful for detecting higher-order properties and informative, relevant patterns, for it yields an immediate description of H and, if the grammar is simultaneously built, a valid η too. However, as Shalizi (2001) notes, “the variables describing emergent properties must be fully determined by lower-level variables.” It becomes clear then that the new simplified “high-level” description is a clever mapping function P of the lower level.

More generally, such methods produce “high-level” description grammars which are still based on an initial lower level (Bonabeau & Dessalles, 1997). In addition, while simpler, the newly created levels are not necessarily (i) more natural and intuitive or (ii) more importantly, complete: their efficiency is indeed limited in case the reductionist approach fails, i.e. when the chosen lower levels are not informative enough about the considered phenomenon.

The question now goes deeper: can an automatic (bottom-up) process yield an *essentially new* vision on things? This sounds as if a deterministic machine could address the problem of ontological uncertainty. In short, it may be hopeless to expect a machine to yield a truly innovative insight starting from already deficient levels. We thus wish to underline that efforts should not necessarily be focused on improving the the design of the model of the dynamics of a given level, using a fixed ontology – although this latter attitude could be encouraged by a reductionist or emergentist stance, as is often the case.

5 Concluding Remarks

On the whole, mistakes are not to be found necessarily in λ , η nor in putative projection functions; but rather in the definition itself of levels L and H . In front of unsuccessful models and simulations, we hence suggest that reductionist and emergentist attitudes in designing models and appraising simulation failures may make it harder to detect ill-conceived modeling ontology and subsequent epistemological dead-ends: some high-level phenomena cannot be explained and reconstructed without a fundamental *viewpoint change* in not only low-level dynamics but also in the design of low-level objects themselves — e.g. introducing *new* glial cells or *new* semantic items, artifacts. In other words, a successful reconstruction may require not only to find a valid and efficient grammar (λ & L), but also to rethink the very bricks that constitute any potential grammar. As suggested above, social and neural network models, at the minimum, could benefit from such hindsight.

Acknowledgements

I wish to thank Michel Bitbol, Paul Bourguine, David Lane, Claes Andersson, David Chavalarias for very fruitful discussions and suggestions. Credit also goes to four anonymous referees and Yannick Kalantzis for interesting feedback. This work has been partially supported by the CNRS. Additionally, some meetings of the EU-funded project “ISCOM” helped framing this paper.

References

1. Anderson, P.: More is different. *Science* 177, 393–396 (1972)
2. Bedau, M.A.: Weak emergence. In: Tomberlin, J. (ed.) *Philosophical perspectives: Mind, causation, and world*, vol. 11, pp. 375–399. Malden, Blackwell (1997)
3. Bitbol, M.: Ontology, matter and emergence. *Phenomenology and the cognitive science* (to appear, 2006)
4. Bonabeau, E., Dessalles, J.-L.: Detection and emergence. *Intellectica* 25(2), 85–94 (1997)
5. Campbell, D.T.: Downward causation. In: Ayala, F., Dobzhansky, T. (eds.) *Hierarchically Organized Biological Systems. Studies in the philosophy of biology*, pp. 179–186. Macmillan Press, Basingstoke (1974)
6. Clark, A.: Emergence and Explanation. In: *Being there: Putting brain, body, and world together again*, ch. 6, pp. 103–128. MIT Press, Cambridge (1996)
7. Crutchfield, J.P.: The calculi of emergence: Computation, dynamics, and induction. *Physica D* 75, 11–54 (1994)
8. Deroian, F.: Formation of social networks and diffusion of innovations. *Research policy* 31, 835–846 (2002)
9. Destexhe, A., Marder, E.: Plasticity in single neuron and circuit computations. *Nature* 431, 789–795 (2004)
10. Emmeche, C., Koppe, S., Stjernfelt, F.: Levels, emergence, and three versions of downward causation. In: Andersen, P.B., Emmeche, C., Finnemann, N.O., Christiansen, P.V. (eds.) *Downward causation. Minds, bodies and matter*, pp. 13–34. Aarhus University Press, Aarhus (2000)

11. Epstein, J.M.: Remarks on the foundations of agent-based generative social science. Tech. rept. 00506024. Santa Fe Institute (2005)
12. Gershenson, C., Heylighen, F.: When can we call a system self-organizing? In: Banzhaf, W., Ziegler, J., Christaller, T., Dittrich, P., Kim, J.T. (eds.) ECAL 2003. LNCS (LNAI), vol. 2801, pp. 606–614. Springer, Heidelberg (2003)
13. Humphreys, P.: Emergence, not supervenience. *Philosophy of science* 64, S337–S345 (1997)
14. Hüttemann, A.: Explanation, emergence, and quantum entanglement. *Philosophy of science* 72(1), 114–127 (2005)
15. Kim, J.: *Mind in a physical world*. MIT Press, Cambridge (1998)
16. Kim, J.: Making sense of emergence. *Philosophical studies* 95, 3–36 (1999)
17. Küppers, G., Lenhard, J.: Validation of simulation: Patterns in the social and natural sciences. *Journal of artificial societies and social simulation* 8(4) (2005), <http://jasss.soc.surrey.ac.uk/8/4/3.html>
18. Lane, D.A.: Hierarchy, complexity, society. In: *Hierarchy in natural and social sciences*. Methodos Series, vol. 3. Springer, Heidelberg (2006)
19. Laughlin, R.B., Pines, D., Schmalian, J., Stojkovic, B.P., Wolynes, P.: The middle way. *PNAS* 97(1), 32–37 (2000)
20. McGregor, S., Fernando, C.: Levels of description: A novel approach to dynamical hierarchies. *Artificial life* 11(4), 459–472 (2005)
21. Nilsson-Jacobi, M.: Hierarchical organization in smooth dynamical systems. *Artificial life* 11(4), 493–512 (2005)
22. Pfrieger, F.W., Barres, B.A.: New views on synapse-glia interactions. *Current opinion in neurobiology* 6, 615–621 (1996)
23. Pumain, D.: Scaling laws and urban systems. SFI Working Paper 04-02-002 (2004)
24. Roth, C.: Empiricism for descriptive social network models. *Physica A* 378(1), 53–58 (2007)
25. Rueger, A.: Robust supervenience and emergence. *Philosophy of science* 67(3), 466–489 (2000)
26. Shalizi, C.R.: Causal architecture, complexity and self-organization in time series and cellular automata, ch. 11. Ph.D. thesis, University of Wisconsin at Madison, U.S.A. (2001)
27. Shalizi, C.R., Shalizi, K.R.: Blind construction of optimal non-linear recursive predictors for discrete sequences. In: Chickering, M., Halpern, J. (eds.) *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pp. 504–511 (2004)
28. Turner, H., Stepney, S.: Rule migration: Exploring a design framework for modelling emergence in CA-like systems. In: *ECAL Workshop on Unconventional Computing (2005)* (to appear in *International Journal of Unconventional Computing*)

Narrative Scenarios, Mediating Formalisms, and the Agent-Based Simulation of Land Use Change

Nicholas M. Gotts and J. Gary Polhill

Macaulay Institute, Aberdeen, Scotland, UK
{n.gotts,g.polhill}@macaulay.ac.uk

Abstract. The kinds of system studied using agent-based simulation are intuitively, and to a considerable extent scientifically, understood through natural language *narrative scenarios*, and that finding systematic and well-founded ways to relate such scenarios to simulation models, and in particular to their outputs, is important in both scientific and policy-related applications of agent-based simulation. The paper outlines a projected approach to the constellation of problems this raises – which derive from the gulf between the semantics of natural and programming languages. It centers on the use of mediating formalisms: ontologies and specialised formalisms for qualitative representation and reasoning. Examples are derived primarily from ongoing work on the simulation of land use change.

Keywords: Narrative scenarios, qualitative ontologies semantics simulation.

1 Introduction

The paper sketches an ambitious programme of work currently at an early stage. It argues that simulation in general, and agent-based simulation in particular, can benefit from the development and use of *mediating formalisms*, to link natural language texts on the one hand, with the program code and input-output behaviour of computational simulation models on the other. These formalisms should have strong qualitative aspects, be computer-readable, and, where possible, support computationally tractable reasoning. The paper concentrates on the role of such formalisms in relation to the potential uses of *narrative scenarios* in building spatially explicit agent-based models of land use change. A narrative scenario, in the sense used here, is a description of the past, or a possible future, with a strong temporal component: a story about what has happened or might happen. Such scenarios may be produced by groups of experts or stakeholders, then used as an input to simulation modelling; or conversely, they may be produced from simulation output, then compared with descriptions of real-world sequences of events, and/or presented to experts or stakeholders for judgments of plausibility.

Computer simulation models are, in general, not readily accessible. Even if the code is available, and well-commented, it is difficult and time-consuming for anyone other than the programmer to understand it. If design documents are not available, it is also necessary to reverse-engineer from the program to the design, to link the code with the high-level description of the model given in any source text (such as a journal paper).

Even if design documents are available, the source code may not necessarily implement the design as expected, as illustrated by Edmonds and Hales's (2003) exploration of Riolo et al.'s (2001) model.

There are several contexts in which it is desirable to link such a model, its input parameters, or its output, to natural language texts, and this raises additional problems. Issues arising when these texts are academic publications, or interview transcripts, are explored in Polhill and Ziervogel (2006), and Polhill and Gotts (2006), and here we consider issues in relation to narrative scenarios, but the main source of the difficulties is that computer programs are written in formal languages, and expressions in natural and formal languages acquire meanings in very different ways. If expressions in a formal language are assigned meanings, this is done by specifying a formal semantics, within which elementary terms are given precisely specified meanings, and complex expressions' meanings are defined using rules for combining the meanings of these elementary terms. In the case of programming languages, the formal semantics will refer to mathematical objects such as partial functions from input to output (denotational semantics: Stoy 1977) or sequences of computational steps (operational semantics: Plotkin 1981) – in any case, the terms of the language refer to nothing beyond the computational domain. Natural language expressions, by contrast, do not have formal semantics, acquiring their meanings from their use in real-world contexts. Expressions within natural languages do not generally have precise meanings, although precision can be increased if desired (if I say “I saw a tall man near the park”, my interlocutor may ask “How tall?” or “Where exactly?” for example). Furthermore, natural language is far more expressive than any formal language. The claim made here is that intermediate languages, formal but designed to refer to and support reasoning about real-world things and processes, can help bridge the natural language – programming language gap.

2 Narrative Scenarios and Historical Social Sciences

The kinds of system we study using social simulation are intuitively (and to a considerable extent scientifically) understood using narratives: natural-language stories about what has happened, might happen, or might have happened, almost always including attributions of causality as an important aspect. Such narratives can be divided into a number of classes, of which the following are most relevant here:

- a) *Historical narratives*: accounts of what has actually happened over some period in the past, within some geographical area and domain of human activity, or to some person or group of people. We are concerned here with narratives intended by their authors to be accurate, although any such narrative is bound to be selective.
- b) *Possible future narratives*: narratives describing what might happen in future. Such narratives may be produced to help prepare for future contingencies, or to justify current policies or argue for particular future ones.
- c) *Counterfactual narratives* or “alternative histories”. Such narratives, when not intended purely for entertainment, may be intended to illuminate the causal structure of what actually happened (e.g. Schmalberger 1998).

- d) *Simulation narratives* derived from simulations or games. Agent-based models, and computer games such as those used in military strategy training, produce sequences of events which can either be considered as occurring within a computational context – as steps in the implementation of an interactive program – or treated as the basis of possible future or counterfactual narratives about the real world.

Historical narratives are an important aspect of what might be called historical social sciences, which include economic history, aspects of social and political history and political science, archeology, paleoanthropology, and historical geography among others. The aims of the historical social sciences include the following, to all of which narrative scenarios are centrally relevant:

- a) *Reconstructing events and sequences of events.* What constitutes an “event”, and how we know whether an event occurred, varies from case to case. While “James McDonald sold Danesbridge Farm to John Robertson for £250,000 on 14th August 1997”¹ is a straightforward case, and ways of verifying or falsifying it are clear (seeking documentary evidence, interviewing the participants and other witnesses), “McDonald became convinced there was no future in farming during the mid-1990s”, “The price of Scottish farmland rose between 1990 and 2000” and “Many Scottish farmers considered leaving farming because of the 2003 CAP reform” are, for different reasons, much less straightforward. In the first of the three, the event concerned is an apparently private one for which we have to rely on McDonald’s memory – although we might seek evidence of statements he made or actions he took – and its temporal location is vague. In the second, the “event” is a summary description of a large set of events (sales of farmland); we might question whether it is an “event” at all, and if it is, what information we need about individual sales to assess its truth: for example, what if records of some sales are missing? The third case combines the difficulties arising in the first two, and adds additional vagueness in its use of “Many”. Turning from single events to sequences of events, narrative scenarios become indispensable to understanding and to further investigation: without the temporal and causal structure that a narrative provides, all we have is a set of unconnected happenings.
- b) *Explaining particular events.* Explanation in the historical social sciences rarely takes the form, frequently found in the experimental physical sciences, of specifying a set of initial conditions and a presumed natural law which ensures that given those initial conditions, the event to be explained must occur (there may be rather trivial exceptions: e.g. explaining the death of King Charles I of England and Scotland, given the initial conditions that he was a human being and that he was decapitated). Rather, an (implicit or explicit) range of possibilities is considered, and causal factors are specified in the absence of which the event being explained would have been impossible or less likely, and some alternative(s) certain or more likely. For example, we might explain Charles losing the Civil War (as opposed to his

¹ The example is an invented one.

- winning it, or a negotiated settlement or stalemate occurring) in terms of his personal qualities, those of his enemies, the balance of ideological and/or class forces in England, military innovations, interactions between English, Scottish and Irish politics, or some combination of these.
- c) *Discovering regularities across time and space.* Two of the “events” referred to above could also be described as regularities: a rise in the price of farmland across Scotland, and the deliberations of “many” farmers following CAP reform. Other regularities, or patterns of events, are more complex, taking the form of correlations between variables, e.g. between agricultural subsidy levels and the price of farmland; or of spatio-temporal patterns of change, such as the spread of agricultural innovations. Some cases of correlations between variables can be investigated without involving narrative scenarios: those which occur at one time but across space, for instance. However, describing co-variation across time at a single spatial location, or co-variation patterns involving both space and time, necessarily involves a narrative element.
- d) *Explaining regularities across time and space.* Sometimes the explanation of a regularity may take something like a law-and-initial-conditions form. For example, when the average price of a commodity rises, demand for it generally falls, while supply (perhaps after a lag) rises. Specific instances of such patterns of events (which are themselves regularities, as they necessarily involve multiple attempts to buy and sell) may be explained by reference to this economic “law”. However, such “laws” do not always hold: for example, a price rise may be taken as a signal that further rises will occur, boosting demand. Frequently, explaining social or historical regularities involves specifying a mechanism or principle claimed to underlie instances of the regularity, that will produce such instances *if nothing else interferes*. This can also be said of most laws in the physical sciences – but there, the experimental method can often be used to satisfy this implicit condition.
- e) *Explaining possibilities:* how could entities/events of type X exist/happen? A central theme in recent social (and biological) science is the explanation of altruistic behaviour (see Gotts, Polhill and Law 2003 for a review), given the obvious advantages of selfishness (in economic and genetic terms). There are in fact a number of plausible candidate explanations, with controversy continuing about which of them contribute.

The last three classes of aim listed indicate how and why the historical social sciences depend on the comparative method: finding events or groups of events which share important features, but differ in one or more crucial respects; and which illuminate regularities across space and time, and patterns of causal influence. If we wish to investigate the rising Scottish farmland prices in the 1990s, for example, we already have a class of events with many features in common. We might then divide them into subclasses (by size or location of farm, date of sale, main agricultural products, age of farmer), and also compare the change in 1990s Scotland prices with changes in England or Wales over the same period, and in Scotland over the 1980s and 1970s, to explain the observed regularity, fit it into wider patterns, and redescribe it in more

illuminating ways. While maps, diagrams and tables can be important in presenting the results of such investigations, none of these will be intelligible without a connecting narrative scenario.

As well as asking whether a historical narrative scenario is accurate – in its account of the course of events, and in causal attribution – we can also ask whether it is *adequate*, in the sense of mentioning the most important events and causal connections. Adequacy must be assessed relative to the narrative's length and function, but given these, the adequacy of two accurate narratives of the same event sequence might be compared by asking whether either fits only of a proper subset of those event sequences the other could truly describe: adequacy thus defines a partial order on accurate historical narrative scenarios.

Turning to possible future scenarios, these are more relevant to policy development than to scientific investigation. Questions about whether they are accurate cannot be answered until the time they refer to, and they can be of considerable use even if the events they describe never happen. However, judgments of *plausibility* are intrinsic to their use in policy development: only narratives policy professionals or stakeholders judge plausible are likely to help them prepare for the future. Judgments of plausibility may be made by asking a range of experts or stakeholders whether they consider a scenario plausible – and if not, what parts or aspects of it are implausible, and why; or by examining it in the light of particular theories of social and historical processes.

Counterfactual narratives are used in historical social sciences, notably in political and military history (Schmalberger 1998), epidemiology (Kay, Prüss and Corvalan 2000) and macroeconomics (Cooper 2004); but many of the same questions concerning plausibility, and the same range of possibilities for assessing this, arise as for possible future scenarios. However, counterfactual scenarios also have an important role in assessing the attributions of causality in historical narrative scenarios: if a causal attribution is valid, then changing a factor to which an important causal role is attributed in bringing about some event should lead plausibly to a scenario in which that event does not occur.

Narrative scenarios derived from computer simulations or games can have important roles in assessing and improving the accuracy and adequacy of historical narrative scenarios, and in assessing the plausibility of possible future and counterfactual narrative scenarios. With regard to historical narrative scenarios, while evidence from the real world will always be the final arbiter of descriptions of the course of events and of regularities, and of proposed causal explanations, simulations have already been used both to test the adequacy of proposed explanations (Lansing and Kremer 1994), and to direct the search for new evidence (Dean et al 1999). We sketch below a systematic approach which fully recognises the role of historical narrative scenarios in historical social sciences, and makes maximum use of the properties of simulation models. This approach would use two largely independent procedures:

- a) Building a simulation model which can produce simulation narratives as similar as possible to an existing historical narrative, from theoretically and/or empirically grounded causal mechanisms, while minimising the number and complexity of additional assumptions required.
- b) As a step in validating an existing simulation model, select two or more parameter sets which can be interpreted in terms of real-world differences in

initial conditions. Provided runs using each parameter set produce outputs which differ systematically in some respect, aspects of the model can then be tested by finding real-world examples of the different kinds of initial conditions, and checking whether real-world outcomes differ in ways corresponding to the differences between simulation outputs. This approach can be adopted without using narrative scenarios, but doing so could greatly strengthen it: because simulation models produce sequences of events, comparison with real-world sequences provides a rich source of information for assessing the model.

Approaches (a) and (b) can be adapted for use with future and counterfactual scenarios: (a) to test the plausibility of an existing future or counterfactual scenario, and (b) to generate a range of new plausible scenarios.

3 Agent-Based Simulations, Narrative Scenarios, and Mediating Formalisms

In this section, we outline some possible intermediate formalisms.

Within computer science, an ontology has been defined as “a formal, explicit specification of a shared conceptualisation” (Gruber 1993). Ontologies formulate relationships between the meanings of a set of terms, combining taxonomies of concepts with information about relations that may hold between entities belonging to specified elements of the taxonomy (e.g. “woman” and “man” might be immediate sub-concepts of “human being”, with additional information specifying possible biological and familial relationships between instances of these concepts). Relations, as well as concepts, can be given a taxonomy: so “father-of” and “mother-of” would both be sub-relations (specialisations) of “parent-of”. Christley et al. (2004) and Polhill and Gotts (2006) argue that ontologies can mitigate problems caused by the ad-hoc way in which agent-based models are programmed. This paper argues that they can also serve to link simulation models and their output to narrative scenarios.

Useful work with ontologies requires a formalism in which to express them: OWL (Antoniou and van Harmelen 2004) is supported by the semantic web community, is compatible with useful ontology-related software (notably Protégé (<http://protege.stanford.edu/>)), and has a sound logical basis in the well-understood description logic SHIQ (Baader, Horrocks and Sattler 2004). Using automated reasoning software an OWL ontology can be checked for consistency, satisfiability of and equivalences among concepts can be inferred, and entities described can be inferred to be instances of particular concepts.

How do we know what a simulation model, a part of that model, or its input and output, represent? In general, the model is described in natural language, possibly accompanied by diagrams, and this description is accompanied by a description of the real world entity or situation (or type of entity or situation) the model is intended to represent. This again is generally couched in some combination of natural language and diagrams; and in fact the description of the model, and of what it is intended to represent, may not be distinguished. There will always be many relevant aspects of the real world not represented in the model, and conversely, many components of the

software implementing the model that are not intended to represent aspects of the real world, but are needed to produce a working program. To the extent it is specified exactly which aspects of the real world are represented in the model, we can say that a *conceptual model* has been defined. To the extent it is specified exactly which components of the model represent these aspects of the real world, the relationship of this conceptual model to the simulation model has been pinned down.

We contend that both the conceptual model, and its relationship to the simulation model, can be greatly clarified by the use of ontologies. This does *not* require the existence of an ontology common to the social simulation community (let alone to social scientists in general): the point is for the authors of a particular simulation model or set of models to share *with* wider communities a precise specification of what they intend their model (and its inputs and outputs) to represent, and which parts of the computer code concerned have representational significance. In fact, a number of ontologies are needed: our ideas on this are still evolving, but figure 1 illustrates one possible setup, showing four types of ontology.

The right of the figure shows a scenario ontology, represented by the darkest and frontmost oval (with other scenario ontologies indicated behind it). The concepts and relations in a scenario ontology represent the types of entity (and individual entities) existing in a part of the real world – or of a possible world as envisaged in a possible future or counterfactual narrative scenario. The scenario may, but need not, be described by one or more natural language narratives of the kind discussed above.

The left of the figure shows a model ontology, again with others of the same kind indicated behind it. The concepts and relations in a model ontology represent the types of *software* entity (and individual entities) that exist in a simulation model: instances of the concepts in a model ontology will be pieces of code or of data, depending on the concept.

At the top of the figure is the domain structure ontology. This is intended to capture what is common to the structure of a set of real-world scenarios and a set of models. Each concept in a scenario ontology will be a sub-concept (specialisation) of some concept in the domain structure ontology, as will each concept in a model ontology. Each model ontology or scenario ontology will “import” the domain structure ontology (copy its structure and contents), then add its own sub-concepts and sub-relations.

The bottom of the figure contains the representation ontology. This imports one scenario ontology and one model ontology, and has the sole purpose of defining the relationship between the two, linking real-world concepts and instances with those representing them in the model. It will include two additional concepts, heading the taxonomies of real-world and software entities.

To give a more concrete idea of what constructing ontologies involves, we include here a draft of the upper layers of a domain structure ontology for use in the land use domain – that is, in relating real-world land use scenarios to agent-based models of land use. The concepts in the ontology are described in real-world terms, but are intended to cover both real-world entities, and the software entities corresponding to them in an agent-based model. The lowest-level entries are *examples* of (relatively) low-level *concepts* (not instances). The sub-concepts below any specific concept are not taken to be exhaustive.

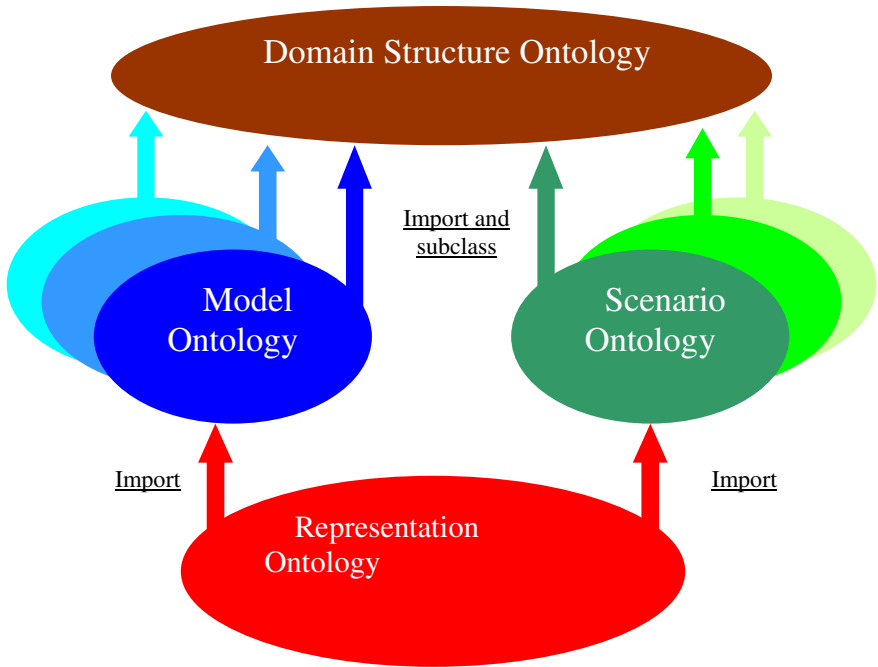


Fig. 1. Ontology relationships

Particular [The top-level node: anything that cannot itself have instances.]

Endurant [Entities that “are ‘in time’, they are ‘wholly present’ (all their proper parts are present) at any time of their existence.” (Masolo, Borgo et al 2003). They are “things” rather than “processes”, contrasting with “Perdurants” - see below.]

PhysicalEndurant

AmountOfMatter [For example, a tonne of grain, a litre of water.]

PhysicalThing

PhysicalObject [Something you could pick up and throw - if you were the right size.]

LivingThing

NonHumanOrganism

DomesticAnimal

Sheep,

Cow,

CropPlant

CornPlant,
BeanPlant,

HumanBeing

Adult,
FemaleHumanBeing

NonLivingPhysicalObject

Tractor, Fence,

FieldOfCrop [Taken to be a
conglomeration of PhysicalObjects and
AmountsOfMatter.]

FieldOfWheat, FieldOfTomatoes

HerdOfAnimals [A collection of
DomesticAnimals. A herd can remain "the
same herd" even while individual
animals come and go.]

HerdOfSheep, HerdOfCows,

Feature [A "Feature" is dependent on a
specific PhysicalThing, to which it
"belongs".]

GapInFence, SkinOfCow,

NonPhysicalEndurant**MentalThing**

Memory, Attitude,

SocialThing**ImpersonalSocialThing**

Law, Currency,

SocialFormation

EthnicGroup, Class,

SocialNetwork

KinshipNetwork,
FriendshipNetwork...

Organisation

FarmersUnion, Corporation,

SocialRole**FormalSocialRole**

Spouse, Landlord,

InformalSocialRole

Friend, RoleModel,

Agent [This subsumes HumanBeing and
Organisation, producing the only examples of
multiple inheritance in this concept
hierarchy.]

Perdurant [States, events, activities, processes -
things which have proper parts in different
temporal locations. All Perdurants involve
Endurants as "participants". All Endurants
participate in Perdurants.]

AgentivePerdurant [A Perdurant involving intentionality on the part of at least one Agent.]

CourseOfAction

LandUse [Apply a land use to a specific LandParcel or LandParcels.], Irrigate,

OneOffAction

Sell, Buy, Ask, Tell,

NonAgentivePerdurant

OnGoingOccurrence

IllHealthBout, Infestation,

OneOffOccurrence

PriceChange, Death,

Location [Endurants and Perdurants "occupy" Locations, and Locations are necessarily defined in relation to Endurants and/or Perdurants.]

SpatialLocation

Province, LandParcel, LandHolding,

TemporalLocation

Year, January,

Spatio-TemporalLocation

Abstract

NetworkThing [Both PhysicalThings and SocialNetworks have a network structure. However, they can change their topology while remaining the same network - which a network considered as a mathematical structure cannot: the current mathematical structure of a real-world Network is thus regarded as a relationship it has with that mathematical structure.]

Network

Tree, DirectedNetwork, UndirectedNetwork,

Node

Link

DirectedLink, UndirectedLink, LabelledLink

Clique

Procedure [A CourseOfAction may involve following a Procedure (as written down, or encoded in memory).]

LandUseProcedure, VeterinaryProcedure,

Ontologies can be used to describe processes and spatio-temporal relationships, but capturing the rich structure of human-produced narrative scenarios, or the output of agent-based simulation models (particularly where these are spatially explicit), is

likely to require additional, specialised formalisms. The most promising include Allen’s temporal interval calculus (Allen and Kautz 1985), the Region Connection Calculus (RCC) (Cohn et al 1997), and Qualitative Differential Equations (QDE) (Kuipers 2001). These are all primarily qualitative, an important advantage in mediating between natural language narratives (which almost always have important qualitative aspects) and simulation output.

Allen’s temporal interval calculus is based on a set of 13 mutually exclusive and jointly exhaustive qualitative relations which two temporal intervals can have (figure 2). Partial knowledge about which of them holds between an interval pair can be expressed by specifying a subset of the 13 asserted to include the true relation. RCC is a somewhat similar set of topological spatial relations. Again, any two regions (which can be of any dimensionality, provided both are of the same dimensionality) must have one and only one of these relations; but in this case, the regions can consist of multiple pieces. The names of the eight relations as given in figure 3 are abbreviations standing for equal (EQ), externally connected (EC), disconnected (DC), partial overlap (PO), tangential proper part (TPP) and non-tangential proper part (NTPP), plus inverses of TPP and NTPP (TPPi and NTPPi).

Both Allen’s calculus and RCC have been extensively investigated with regard to their computational properties. For RCC, the full first-order language is undecidable (Grzegorzczuk 1951, Gotts 1996); it would be surprising if the situation were different with Allen’s calculus, although its first-order language does not appear to have been investigated explicitly. For both, the “constraint language” is decidable – that is, given any finite set of relations among regions, it can be determined in finite time whether it could be satisfied; but the problem is NP-hard (Vilain, Kautz and van Beek 1990, Bennett 1994, 1996). For both calculi, subsets of this class of problems are known for which difficulty can be shown to increase with some polynomial function of problem

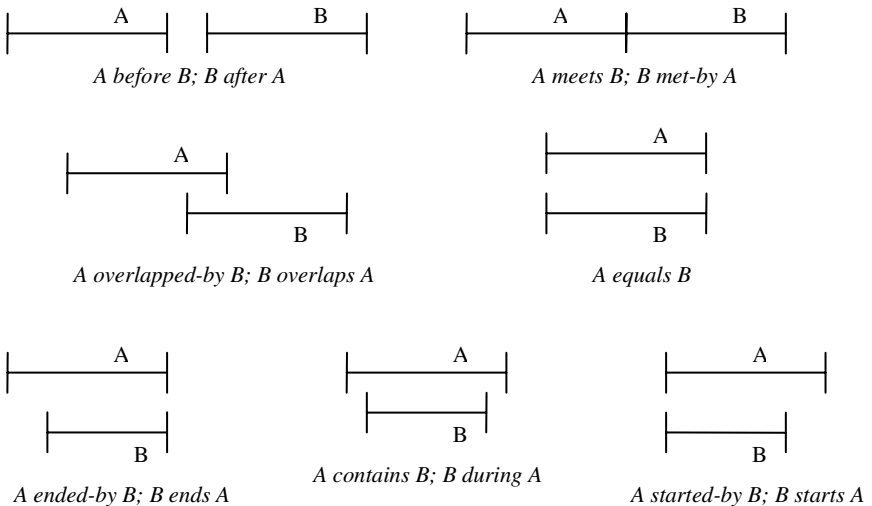


Fig. 2. Allen’s 13 qualitative temporal interval relations

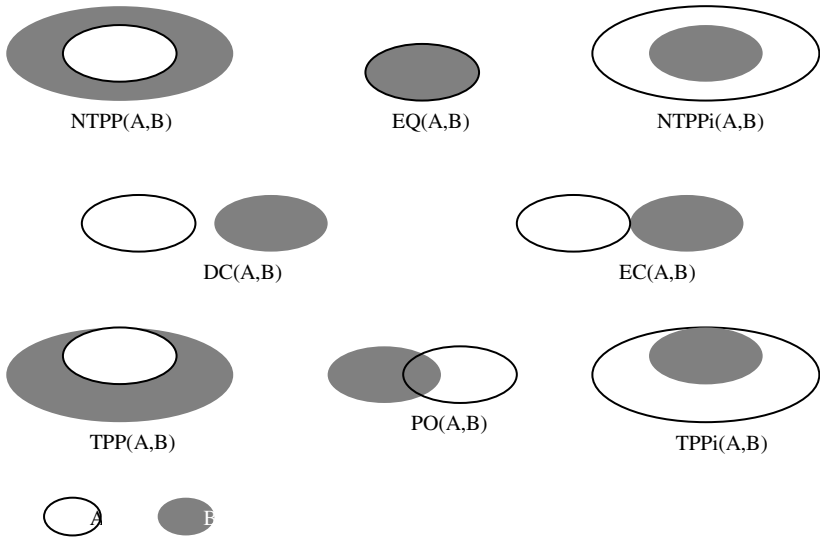


Fig. 3. RCC's eight basic topological relations

size (Vilain, Kautz and van Beek 1990, Bennett 1994, 1996, Renz and Nebel 1999). Bennett et al (2002) show that RCC and Allen's calculus can be combined to express spatial and temporal relations simultaneously without losing decidability.

Both Allen's calculus and RCC leave large gaps in what can be said about spatial and temporal properties and relations (they have no metric component, and RCC as presented above says nothing about shape, although there are extensions allowing the convexity of a region to be asserted (Davis, Gotts and Cohn 1999) and for the existence of regions without definite boundaries (Cohn et al 1997)). However, what they can express is highly relevant to questions of causality: in the everyday world, a cause must not come after its effect, and the two must be spatio-temporally connected (directly or via some intermediate). Qualitative temporal relationships are also central to the description of plans and procedures, and concerns about boundaries and the spatial continuity of parcels of land are highly relevant to land management, and to the ecological implications of land use change. Given this, and the extensive literature about the expressivity and computational properties of the two formalisms, we intend to investigate combining them with ontologies in representing narrative scenarios.

In a QDE representation of a system, each numerical variable is assigned a *quantity space*: a finite, totally ordered set of qualitatively important "landmark" values. In a land use context, taking rainfall over the growing season as an example, the landmark values might be those (not necessarily specified exactly) necessary to make various crops viable. Variables can be related by algebraic constraints (e.g., the return from a crop is the multiple of its yield with the price per unit weight), or by differential ones (e.g. the relation of the rate of pollutant inflow to a closed body of water to the quantity of that pollutant in the lake); functional constraints assert that increasing one variable will increase (or decrease) another, without further specifying the form of the dependence. Transition conditions may be attached to a set of QDE constraints, specifying

when they cease to apply (e.g. increasing the number of sheep in a field will increase their total rate of weight gain, but only up to the point where they eat the grass faster than it can grow).

4 From Narrative Scenarios to Simulations...and Back

We sketch here a development of the proposed “story and simulation” approach to scenario development (European Environment Agency 2001):

“The storyline describes in story form how relevant events unfold in the future, while the model calculations complement the storyline by presenting numerical estimates of future environmental indicators and helping to maintain the consistency of the storyline.”

A small set of storylines, each based on different assumptions, is developed by a “scenario panel” of experts and/or stakeholders. A modelling team then creates a simulation to match each storyline; the additional (quantitative) detail and any caveats about consistency are fed back to the scenario panel, this process being repeated as necessary. However, the process of getting from storyline to model is not described in any detail, nor is the kind of feedback given. We suggest the following sequence of steps, beginning with a set of natural language possible future narrative scenarios, an initial domain structure ontology, qualitative formalisms such as those described in the preceding section, and possibly a set of existing scenario ontologies. All the steps listed require further decomposition. We assume that the domain structure ontology includes at least the concepts in the taxonomy in section 3:

1. Identify concepts in the narrative scenarios possibly relevant to simulation modelling, and instances of those concepts. Natural language processing software such as GATE (Cunningham, Maynard et al 2002), which can pick out words and phrases using syntactic and semantic criteria, may be useful here and in later stages, but full automation is a long way from feasibility.
 - a. Identify the Agents (HumanBeings or SocialFormations) in the narratives.
 - b. Identify the non-agentive Endurants (Physical and NonPhysical).
 - c. Identify the AgentivePerdurants, and the Endurants involved in each.
 - d. Identify the NonAgentivePerdurants, and the Endurants involved in each.
 - e. For each Perdurant and Endurant, identify any Locations mentioned in connection with it. There may be SpatialLocations, TemporalLocations and/or SpatioTemporalLocations specifying where and when the Perdurant happened.
2. Identify relations between instances of the concepts identified in step 1, and properties of those instances. Particularly important will be relations between the Locations identified in step 1e: existing systematisations of such relations, such as those described above, should be part of the domain structure ontology.

3. For each concept identified in step 1, find the most specific concept in the domain structure ontology of which it can be considered a sub-concept.
4. Where more than one of the concepts identified in step 1 has the same immediate super-concept in the domain structure ontology, consider whether any of that set of concepts should be grouped under more specific, intermediate concepts. If so, add these to the scenario ontology.
5. Carry out analogues of steps 3-4 for relations and properties.
6. On the basis of commonsense knowledge, stakeholder knowledge and/or existing theoretical and empirical literature, consider what additional real-world concepts, properties and relations should be added to the scenario ontology.
7. Construct a formal description of *each* narrative scenario, in terms of the scenario ontology and the set of qualitative formalisms being used. This formal description would take the form a labelled, directed graph. Details are still under consideration, but provisionally:
 - a. Assign a node to each of the specific Endurants, Perdurants and Locations mentioned in that narrative scenario or inferred by combining it with background knowledge, labelled with the concepts of which it is an instance.
 - b. Link each Perdurant to all the Endurants involved in it, with a label on the link identifying the role played by that Endurant.
 - c. Link each Perdurant to the TemporalLocation it occupied, and to one or more SpatialLocations (one of these would be the spatial union of all the SpatialLocations involved; further details remain to be decided).
 - d. Instances of relations between Endurants, between Perdurants, between Endurants or Perdurants and Abstracts, and relations between an Endurant and a Perdurant other than that of participation (e.g., a HumanBeing *learning* of a Perdurant) should also have nodes, with edges linking them to each of the concept-instances involved in the relation-instance, and an edge linking the relation instance to the TemporalLocation during which it held (one such node would represent the “universal” TemporalLocation, indicating that a Perdurant or relation linked to it continued or held throughout the time covered by the scenario).
 - e. Properties of Endurants and Perdurants should also have nodes, with a link to one Endurant or Perdurant, and to a set of TemporalLocations, the latter labelled with the values holding during that TemporalLocation (these could be taken from a QDE-type quantity space in the case of numerical properties).
 - f. SpatialLocations could have RCC-labelled links with each other, TemporalLocations interval-relation links with each other.
 - g. Pairs of property-nodes could also be linked to nodes representing QDE algebraic, differential or functional constraints.
8. Decide on a subset of the concepts, relations and properties identified to be represented in the simulation model, and determine how to represent them.

9. On the basis of step 8, construct a model ontology. A concept in the scenario ontology is likely to have relations and properties which the model ontology concept for the class of software entity representing that real world entity does not have, and vice versa. Polhill and Gotts (2006) discusses the information to be encoded in the concepts of a model ontology.
10. For each pair of corresponding concepts (relations, properties) in the scenario and model ontology, consider whether a new concept (relation, property) subsuming both (and nothing else) should be added to the domain structure ontology. This should be done if and only if two concepts (relations, properties) share structural relationships with other items in their respective ontologies which do not hold for their current common super-concept in the domain structure ontology.
11. Build the model. This process may modify an existing model, may require a new model to be built within an existing modelling system (see Polhill and Gotts (2006) for discussion of modelling systems), or may require an entirely new modelling system.
12. Experiment with the model, exploring its parameter space to discover whether simulation runs giving rise to simulation scenarios similar to the members of the original set of narrative scenarios can be produced, and if so, how easily, and where the greatest difficulties lie.
13. Feed back information from step 12 to the experts/stakeholders (along with a description of the simulation model), and if there were indeed significant difficulties in producing simulation narratives similar to the original narrative scenarios, ask whether they can adjust these accordingly.

Given a simulation model based on an ontology, it should be possible to arrange for output from a simulation run to take the form of a directed, labelled graph such as that described in step 8 above. Matching such structures against each other should also be possible, using adaptations of existing graph-matching algorithms such as that described in Feng, Goldstone and Menkov (2004), which is designed for aligning conceptual systems, and allows for post-matching adjustment by the user. This would allow systematic comparison with the output from a different run of the same model; with outputs from different models, and with narrative scenario structures produced from natural language narratives, and perhaps other sources such as time series. We intend to explore how a natural language narrative scenario could be created from such a labelled, directed graph – specifically, whether there are existing natural language generation algorithms that could be adapted to partially automate the process.

5 Conclusions

Narrative scenarios play crucial roles in both intuitive and scientific understanding of the kinds of system agent-based social simulation models. Indeed, this fact is closely related to the advantages of agent-based models in studying particular types of system: where we conceive of a system as consisting of a set of interacting agents, each with a characteristic range of behaviors (and perhaps perceptions, abilities and goals); then an agent-based model can mirror the structure of the system as we conceive and describe it far more readily than a model based on a system of differential equations.

As Chattoe (1996) argues, this gives agent-based modelling considerable methodological advantages, in particular making it easier to identify the part of the model responsible for poor performance. Such systems are also among those about which a great deal can usefully be said in qualitative terms, even if little quantitative information is available (although not unique in that regard).

Since agent-based models naturally produce outputs with a quasi-temporal structure, and spatially explicit models produce outputs with quasi-spatial structure, ways must be found to match such outputs against natural language narrative scenarios, if such models are to fulfill their potential in historical social sciences and in policy development. However, the transparency which both scientific and policy-related applications require of agent-based simulation modelling in areas such as land use change, cannot be achieved by any combination of program code and natural language description alone. Ontologies, combined with existing qualitative formalisms designed to express spatial, temporal and dynamical relationships and associated with useful results concerning expressivity and computational complexity, could be of great benefit in this regard, and specifically in enabling simulation model outputs to be linked to and compared with natural language narrative scenarios.

Acknowledgements

Work for this paper was funded by the Scottish Executive Environment and Rural Affairs Department, and by the European Commission as part of NEST Pathfinder CAVES project 12186 (<http://cfpm.org/caves/>).

References

1. Allen, J.F., Kautz, H.: A model of naive temporal reasoning. In: Hobbs, J., Moore, R. (eds.) *Formal Theories of the Commonsense World*, Norwood, NJ, pp. 251–268 (1985)
2. Antoniou, G., van Harmelen, F.: *Web Ontology Language: OWL*. In: Staab, S., Studer, R. (eds.) *Handbook on Ontologies*, ch. 4. Springer, Heidelberg (2004)
3. Baader, F., Horrocks, I., Sattler, U.: *Description logics*. In: Staab, S., Studer, R. (eds.) *Handbook on Ontologies*, ch. 1. Springer, Heidelberg (2004)
4. Bennett, B.: Spatial reasoning with propositional logics. In: Doyle, J., Sandewall, E., Torasso, P. (eds.) *Proceedings of the 4th International conference on Principles of Knowledge Representation and Reasoning (KR 1994)*, pp. 51–62 (1994)
5. Bennett, B.: Modal logics for qualitative spatial reasoning. *Bulletin of the Interest Group in Pure and Applied Logic (IGPL)* 4(1), 23–45 (1996)
6. Bennett, B., Cohn, A.G., Wolter, F., Zakharyashev, M.: Multi-dimensional modal logic as a framework for spatio-temporal reasoning. *Applied Intelligence* 17(3), 239–251 (2002)
7. Chattoe, E.: Why are we simulating anyway? In: Troitzsch, K.G., Mueller, U., Gilbert, N., Doran, J.E. (eds.) *Some answers from economics. Social Science Microsimulation*, pp. 78–104. Springer, Heidelberg (1996)
8. Christley, S., Xiang, X., Madey, G.: Ontology for agent-based modeling and simulation. In: Macal, C.M., Sallach, D., North, M.J. (eds.) *Proceedings of the Agent 2004 Conference on Social Dynamics: Interaction, Reflexivity and Emergence*, co-sponsored by Argonne National Laboratory and The University of Chicago, October 7-9 (2004), <http://www.agent2005.anl.gov/Agent2004.pdf>

9. Cohn, A.G., Bennett, B., Gooday, J., Gotts, N.M.: RCC: a calculus for region based qualitative spatial reasoning. *Geoinformatica* 1, 275–316 (1997)
10. Cooper, A.: The impact of interest rates and the housing market on the UK economy. *Economic Outlook* 28, 10–18 (2004)
11. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In: Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL 2002), Philadelphia (July 2002)
12. Davis, E., Gotts, N.M., Cohn, A.G.: Constraint networks of topological relations and complexity. *Constraints* 4, 241–280 (1999)
13. Dean, J.S., Gumerman, G.J., Epstein, J.M., Axtell, R.L., Swedlund, A.C., Parker, M.T., McCarroll, S.: Understanding Anasazi culture change through agent-based modelling. In: Kohler, T., Gumerman, G. (eds.) *Dynamics in Human and Primate Societies*, pp. 179–205. Oxford University Press, Oxford (1999)
14. Edmonds, B., Hales, D.: Replication, replication and replication: Some hard lessons from model alignment. *Journal of Artificial Societies and Social Simulation* 6(4), <http://jasss.soc.surrey.ac.uk/6/4/11.html>
15. European Environment Agency, Scenarios as Tools for International Environmental Assessments, Environmental issue report No. 24, Luxembourg Office for Official Publications of the European Communities (2001)
16. Feng, Y., Goldstone, R., Menkov, V.: A graph matching algorithm and its application to conceptual system translation. *International Journal on Artificial Intelligence Tools* 14(1-2), 77–99 (2004)
17. Gotts, N.M.: An axiomatic approach to topology for spatial information systems. Technical report 96.25, School of Computer Studies, University of Leeds (1996)
18. Gruber, T.R.: A translation approach to portable ontology specifications. *Knowledge Acquisition* 5, 199–220 (1993)
19. Grzegorzczak, A.: Undecidability of some topological theories. *Fundamenta Mathematicae* 38, 137–152 (1951)
20. Kay, D., Prüss, A., Corvalan, C.: Methodology for assessment of environmental burden of disease: Report on the ISEE session on environmental burden of disease, Buffalo. WHO/SDE/WSH/00.7 (August 22, 2000), <http://www.who.int/docstore/peh/burden/wsh00-7/Methodan1.htm>
21. Kuipers, B.: Qualitative Simulation. In: Meyers, R.A. (ed.) *Encyclopedia of Physical Science and Technology*, 3rd edn., pp. 287–300 (2001)
22. Lansing, J.S., Kremer, J.N.: Emergent properties of Balinese water temple networks: coadaptation on a rugged fitness landscape. In: Langton, C.G. (ed.) *Artificial life III*, pp. 201–223. Addison-Wesley, Reading (1994)
23. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., Schneider, L.: (2003)
24. The WonderWeb Library of Foundational Ontologies Final Report WonderWeb Deliverable D18, version 1.0, IST Programme of the Commission of the European Communities project number IST-2001-33052
25. Plotkin, G.D.: A Structural Approach to Operational Semantics, Tech. Rep. DAIMI FN-19, Computer Science Department, Aarhus University, Aarhus, Denmark (1981)
26. Polhill, J.G., Gotts, N.M.: A new approach to modelling frameworks. In: First World Conference on Social Simulation (WCSS 2006), Kyoto, Japan, August 21–25 (2006)
27. Polhill, J.G., Ziervogel, G.: Using ontologies with case studies: an end-user perspective on OWL. In: Second National Conference on e-Social Science (NceSS 2006), Manchester, June 29–30 (2006)

28. Randell, D.A., Cui, Z., Cohn, A.G.: A spatial logic based on regions and connection. In: Nebel, B., Swartout, W., Rich, C. (eds.) *Proceedings of Third International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, Massachusetts, pp. 165–176. Morgan Kaufman, San Francisco (1992)
29. Renz, J., Nebel, B.: On the complexity of qualitative spatial reasoning: A maximal tractable fragment of the Region Connection Calculus. *Artificial Intelligence* 108, 69–123 (1999)
30. Riolo, R.L., Cohen, M.D., Axelrod, R.: Evolution of cooperation without reciprocity. *Nature* 411, 441–443 (2001)
31. Schmalberger, T.: *Dangerous Liaisons: A Theory of Threat Relationships in International Politics*. PhD thesis, available online (1998), http://www.usc.edu/dept/LAS/ir/cews/html_pages/Dangerous_Liaisons/
32. Stoy, J.E.: *Denotational Semantics: The Scott-Strachey Approach to Programming Language Semantics*. MIT Press, Cambridge (1977)

Validation and Verification in Social Simulation: Patterns and Clarification of Terminology

Nuno David

ISCTE - Instituto Superior de Ciências do Trabalho e da Empresa

Department of Information Science and Technology

Lisboa, Portugal

Nuno.David@iscte.pt

Abstract. The terms ‘verification’ and ‘validation’ are widely used in science, both in the natural and the social sciences. They are extensively used in simulation, often associated with the need to evaluate models in different stages of the simulation development process. Frequently, terminological ambiguities arise when researchers conflate, along the simulation development process, the technical meanings of both terms with other meanings found in the philosophy of science and the social sciences. This article considers the problem of verification and validation in social science simulation along five perspectives: The reasons to address terminological issues in simulation; the meaning of the terms in the philosophical sense of the problem of “truth”; the observation that some debates about these terms in simulation are inadvertently more terminological than epistemological; the meaning of the terms in the technical context of the simulation development process; and finally, a comprehensive outline of the relation between terminology used in simulation, different types of models used in the development process and different epistemological perspectives.

Keywords: Verification, validation, pre-computerized, post-computerized models, terminology.

1 Introduction: On the Reasons to Address Terminology

Verification and validation are two extensively used terms in simulation. They are widely used in science in general, both in the natural and the social sciences. They have plethora of different methodological significances, in diverse epistemological perspectives, upon different beliefs, and expectations. They are used often with the same or interchangeable meanings. They are the subject of numerous scientific and philosophical debates, and connected to diverse disciplinary, interdisciplinary and multidisciplinary contexts. In spite of recalcitrant debates, a standard meaning is unlikely to emerge. The terms carry an ineluctable relationship to the problem of “truth”. In simulation, their affirmative character, alluding to a positive result claim, is often criticized (e.g. see Oreskes et al. 1994). Terminological disputes seem unlikely to be useful. Consensus in meaning seems improbable. In effect, it would seem reasonable to ask: Would it be useful, or even reasonable, to discuss these terms on strict terminological grounds without discussing their epistemological underpinnings?

There are, however, methodological contexts where the terms acquire an increased pragmatic semantics. While scientific practice tends to lead scientific communities towards using common methods, as well as common languages and meanings, the use of these terms tends to become considerably pragmatic. Particularly in disciplines that make use of computerized models, like in simulation, the terms verification and validation are associated with the need to evaluate models in different stages of the software development process. In social science simulation, common definitions are usually imported from computer science, as well as from numerical and technical simulation, having intended distinct – but not infrequently conflated – meanings. In the strict context of the simulation development process, it would not be unsafe to say that the purposes of verifying and validating a simulation have become reasonably consensual:

Verification – Most researchers define verification as referring to the performance of the program code, for instance, as “the process of checking that a program does what it was planned to do” (Gilbert and Troitzsch, 1999, p.21), or “checking that the representation is faithful to the simulator’s intentions” (Edmonds, 2003, p.108);

Validation – According to Gilbert and Troitzsch (1999, p.22), “While verification concerns whether the program is working as the researcher expects it to, validation concerns whether the simulation is a good model of the target”, or, according to Edmonds (2003, p.108), that “the expression of the simulation in terms of outcomes is faithful to the relevant social phenomena.”

Comparable definitions abound in the literature. At any rate, given the strong epistemological character of both terms it is always possible to perceive differences among similar definitions. Consider for instance the term *validation*. Whereas Gilbert and Troitzsch seem to define it with reference to the process of simulation as a whole, Edmonds emphasises the expression of the simulation in terms of its outcomes. However, as Küppers and Lenhard (2005) have shown, it is not unusual to construct useful simulations with faithful outcomes even when new assumptions that contradict the intended conceptual model are introduced in the specification and the corresponding programs. Whereas the simulation program may not be a good model of the target, the resulting data output model may reflect a good model of it. There is in fact more than one model involved. Could we say that in Gilbert and Troitzsch’s sense the simulation would not be validated, but that it would be so in Edmonds’ sense?

Although useful in specific contexts, terminological debates would not seem useful if the attempt was to confront or falsify definitions of terms on comprehensive contexts. In effect, as it is widely recognized, several chains of intermediate models are developed before obtaining a satisfactory simulation. Moreover, given the multi-paradigmatic character of the social sciences, many kinds of models will be used in different simulations. And this begs the question: Is it useful to discuss patterns of terminology without outlining the specific methodological context of a simulation and its intrinsic epistemological ambiguity?

Epistemological significances are, after all, the fundamental ingredients to outline a logic of the method of simulation.¹ Indeed, the significances of these terms are often

¹ Previous work on the analysis of intentional and empirical adequacy of computer programs in social science simulation is such an example (see David et al. 2005).

built over tacit assumptions, which tend to evolve as a cultural aspect of the discipline, as a function of paradigmatic diversity and change – there are not and there should not be strict rules about it. On the other hand, if the issue is methodological clarity there may be reasons to address a terminological analysis. If the goal is delimiting peculiar concepts for particular terms, relative to particular methods – or, in other words, if the goal is to disambiguate the mapping between pragmatic terminological use and ontological/epistemological significances – there may be room for useful argument. Why would that be interesting? Essentially because in social science simulation there are two distinct senses in which the terms *verification* and *validation* are used, which support the reasons for writing this article: *verification and validation in the computational modelling sense* and *verification, validation and confirmation in the traditional philosophical sense*. I will often call the former VERIFICATION AND VALIDATION IN THE SIMULATION DEVELOPMENT SENSE and the latter VERIFICATION, VALIDATION AND CONFIRMATION IN THE BROAD SENSE.

2 On the Goals and the Structure of This Article

The lack of consensus with respect to the nature of scientific knowledge suggests that the terms represented by ‘verification’ and ‘validation’ will never be completely free from a certain degree of ambiguity. The ambiguity seems somehow more salient in social science simulation insofar as researchers conflate the meanings of both terms along the simulation development process with other meanings found in philosophy of science and the social sciences. The goal of this article is the clarification of terminology. I will suggest that the identification of distinct phases along the construction of a simulation contributes to clarify the use of the terms. Particular emphasis will be given to the importance of recognizing the construction of two distinct kinds of conceptual models in the simulation development process, one before the implementation and execution of the simulation programs – THE PRE-COMPUTARIZED MODEL – and another after the implementation and execution – THE POST-COMPUTARIZED MODEL.

This somehow obvious distinction among pre-computational, post-computational and actual computerised models does not appear to be methodologically relevant in computer science or technical simulation, but its tacitness contributes to increased ambiguity in usage of the terms in social science simulation. The consideration of different kinds of intermediate models in the logic of the method of simulation helps us situate the roles of the terms in the development process and thus into the real epistemological debate. Additionally, the consideration of the meaning of both terms in the development process sense, on the one hand, and in the broad sense, on the other hand, helps us describe the different categories of knowledge involved in building simulations. It will further allow us to observe that many debates around the problem of “truth” in simulation are inadvertently more of a terminological nature than of an epistemological nature.

3 On the Meaning of Verification, Validation and Confirmation of Models in a Broad Sense

The well known paper of Oreskes et al. (1994) is arguably the reference of criticism to an alleged misleading use of terms in simulation. For some researchers, it is actually a

criticism to simulation as method. The paper emphasised the role of numerical models in the earth sciences and its influence on public policy. Today, in the era of global warming, no one would deny the importance of computers and simulation in the natural sciences for public policy. In any case, the claim is their abstract is undoubtedly the following: “Verification and validation of numerical models of natural systems is impossible” (Oreskes et al. 2004, 641). More than ten years after the publication of this article, refutations abound in the literature. In a widely accessed lecture on the internet,² Tetsuji Iseda, a philosopher and former research assistant of Frederick Suppe³, confronts Suppe’s and Oreskes’ ideas on the value of simulation. A debate which, according to Iseda, should converge to a central question: “Can simulation models yield knowledge about the real world?” And the answers of Suppe and Oreskes, in the view of Iseda, are the following:

“Oreskes et al: We can not *verify* simulation models, so scientists cannot obtain knowledge from simulation modelling.”

“Suppe: Simulation models can be *verified* in some sense, so we can obtain knowledge from them.”

The common term at stake is the word “verification”. For Iseda, however, there are various degrees of certainty on the problem of “truth”. These are identified by the terms verification, confirmation and validation, recalled by Oseda in the light of Oreskes’ critique of simulation:

Firstly, VERIFICATION stands for absolute truth, “To say that a model is verified is to say that its truth has been demonstrated, which implies its reliability as a basis for decision-making” (Oreskes et al. 2004, 641). Models cannot be verified insofar as the real world is never a closed system, and so there cannot be a logical proof that a model is true. Closed systems come up only in purely logical structures, such as proofs in formal logic. Secondly, VALIDATION stands for a “model that does not contain known or detectable flaws and is internally consistent” (*ibidem*, 642). Models can be validated by comparing different solutions, or by calibrating models to adjust known data. But this does not mean that models have been verified, insofar as “the agreement between measures and numerical outputs does not demonstrate that the model that produced the output is an accurate representation of the real system” (*ibidem*, 642). Finally, CONFIRMATION stands for plausible theories in terms of evidence. “If empirical observations are framed as deductive consequences of a general theory and these observations can be shown to be true, then the theory is confirmed and remains in contention for truth” (*ibidem*, 643). Hence, as Iseda remarks, models may be confirmed but this means only the model is probable, not that the model is true.

In summary, for Oreskes et al., simulations can be validated and confirmed but never verified. The primary value of simulation would be to “offer evidence to strengthen what may be already partly established, or to offer heuristic guidance as to further research, but never susceptible to proof.” For Oreskes, a simulation is thus more like a

² Iseda (1997), see <http://carnap.umd.edu/phil250/oreskes/oreskes.html> (retrieved March 2007).

³ The author, among others, of the “Semantic Conception” of scientific theories.

fiction: “A model, like a novel may resonate with nature, but it is not a “real” thing” (*ibidem*, 644). Conversely, for Iseda, verification is possible in some sense. A simulation is a representation of aspects of the real world and thus yields knowledge about the real world: Oreskes’ way of defining “verify” would be too strict – “it would make empirical knowledge impossible” (Iseda 1997). Arguably, it may not be unsafe to say that Oreskes’ view is shared by many researchers doing simulation. However, rather than an epistemological debate, this seems to be a terminological discussion. My objection to Oreskes’ interpretation about Oreskes’ essay would be the following:

- (i) Oreskes’ view is not incompatible with Suppe’s; Oreskes’ criticism is mainly on the use of terminology, claimed to be misleading; (ii) Oreskes’ criticism to terminology does not show that simulation models cannot be validated and/or confirmed; (iii) it is true that Oreskes’ sense of the term verification is strict and thus simulations cannot be verified in the sense defined, but whether Oreskes has suggested or not that simulation cannot yield knowledge about the real world is by no means clear in Oreskes’ paper.

Rather than questioning if simulation yields knowledge, there is another question, which has not been addressed by Oreskes or Iseda, which would be perhaps more illuminating: What *kind* of knowledge can simulation yield about the real world?

Not only this question is more pertinent from an epistemological point of view, but we argue to be useful for clarifying the mapping of terminological usage to epistemological significances, avoiding what could be a mere debate around the use of the same terms in distinct contexts. The simulation development process involves the construction of several models, built according to different methods, serving different purposes, ultimately embedded in the simulation and eventually yielding different kinds of knowledge to the researcher. And this brings us back to our central remark: the terms verification, validation and confirmation of models are usually discussed in the broad philosophical sense, but are often conflated with their use in the development process sense. An ambiguity that led Oreskes to state the following contentious sentence about verification:

“Numerical models may contain closed mathematical components that may be verifiable, just as an algorithm within a computer program may be verifiable.” (Oreskes et al., 1994, p.641).

And further suggesting that the following definition of validation, given by the International Atomic Energy Agency, is erroneous:

“A validated model is one that provides a good representation of the actual processes occurring in a real system.” (Oreskes et al., 1994, p.642)

However, outside the traditional sense of computer science, the first quoted sentence may also be erroneous. Whether an algorithm within a computer program is verifiable or not depends on whether the program may be qualified as a pure logical structure or as a causal model of a logical structure that instantiates a particular algorithm (see Fetzer, 1999 and also David et al. 2005). Even the simplest program may

not be verified in Oreskes' sense. Moreover, whereas the second quoted sentence may be misleading when interpreted in the light of the philosophical sense of validation, it would be actually consensual in the strict computational modelling sense of simulation, in which the goal is to evaluate whether the specification and the results of a simulation are good models of the target, which are two well defined steps in the simulation development process. In short, whether validation in the computational modelling sense stands for verification, validation or confirmation in the broad philosophical sense, it may well be a relevant terminological debate, but not necessarily a useful epistemological debate. For someone not familiar with simulation and from an epistemological point of view the terms may indeed lead to confusion, but are not necessarily erroneous.

4 On Four More Examples of Terminological Dilemmas in Social Science Simulation

Isedas' note on Oreskes *vs.* Suppe is mostly related to simulation in the natural sciences. Likewise, not considering terminological and epistemological traditions in the computer and the social sciences may yield misleading methodological analyses of social science simulation. Apart from the conflation of the terms in the development sense with the philosophical broad sense, there are other relevant ambiguities that may have a terminological rather than an epistemological origin. This observation is not a criticism of the methodological literature in the field. It is rather a natural consequence of an intense interdisciplinary field with exciting new methodological challenges. In any case, the clarification of terms and the resultant emergence of epistemological debates is a positive contribution to the field. The following topics are examples liable to terminological dilemmas, with potential influence on epistemological debates – or simply epistemological misunderstandings, mostly arising as a product of the encounter of the computer and the social sciences:

Experimental vs. Empirical vs. Quasi-empirical Methodology. To the extent that simulation is an experimental methodology, it has been considered as a quasi-empirical approach (e.g. Küppers and Lenhard, 2005, paragraph 2.2). Moreover, insofar as it is programs that are executed and tested, and not the phenomena that they presumably represent, the term “quasi-empirical” has replaced the term “empirical.” Both terms used in this sense seem to be borrowed from the natural sciences, and not from the sense of building and evaluating social science simulations with gathered empirical data. If the intention is to indicate that simulation resembles empirical research to some degree, in the sense that gives a quasi-access to a real experiment with the social phenomena, the term may become highly controversial, and indeed may be claimed misleading. Rather than stating that simulation is quasi-empirical, consider the ambiguity in stating that simulation makes “quasi-experiments” with the social phenomena. Moreover, just like the use of hypotheses confirmed by observation does not necessarily imply using experiments, there may be not any reasons to suggest that

simulation, as a definitive experimental exercise with computerized models, resembles empirical science in the sense of experimental science. At the very least, if the term empirical is used to characterize the logic of the method of simulation, its intended meaning should be clarified.

Formal inference in the classical computer theory sense vs. inference in the discursive intentional sense (for more details, see David et al. 2005). Epstein (1999), among others, has used the “formal symbol manipulation” conception of the classic theory of computation, which describes computers as formal inference machines, to characterize generative simulation as necessarily deductive. The terms “generative” is now widely adopted in social science simulation. However, to understand simulation according to such formal conception of computation requires understanding the inference mechanism in the exclusive grounds of first-order logical sentences. But there are not reasons to suppose that the semantically rich social phenomena represented in simulations can be expressed simply with first-order logical sentences. The building and evaluation of simulations involve the use of rhetorical and intentional discourse, usually acquired in the context of limited consensus. While those simulations may be formally described in the sense of being abstract, it should be remarked that the term formal in such a sense does not necessarily translate to the sense of “formal symbol manipulation” utilized in classic computer science.

Deduction in the formal sense vs. deduction in the empirical sense (for more details, see David et al. 2007). While Axelrod characterised simulation as a contrast to both induction and deduction, induction was defined as the “discovery of patterns in empirical data” and deduction as the specification of a “set of axioms and proving consequences that can be derived from assumptions.” Deduction, in this sense, seems to refer to pure mathematical-logical demonstrations, as opposed to deduction as a kind of ampliative reasoning in an empirical sense. Yet, there is no strong reason for not viewing deduction as a kind of empirical enquiry. Popper, like many other deductivists in the philosophy of science, would say that there is no such thing as induction. Indeed, to define the epistemic specificities of simulation based on the contrast between induction and deduction does not seem to be methodologically informative. Whether the epistemic conception of empirical enquiry may be understood as inductive, deductive or even abductive is by no means a specific dilemma of simulation.

Abduction vs. generativism. The notion of “generative” was initially posed as the specific epistemic characteristic of simulation that contrasts with deduction and induction, a sort of new methodological conception of doing science, often described as a “third way of doing science”.⁴ At any rate, the term “generative” seems to be just a synonymous of “abductive”. Indeed, what the term generative has been characterising seems to be similar to Pierces’ second conception of abduction, in which hypothetical explanations are inquired in order to explain a given explanandum. If that is the case, from an epistemological point of view the term “generative” does not seem to describe a new methodological conception in science. Moreover, “generativist” claims may well run the risk of being considered overenthusiastic or even misleading.

⁴ See Epstein (1999). See also Axelrod (1997).

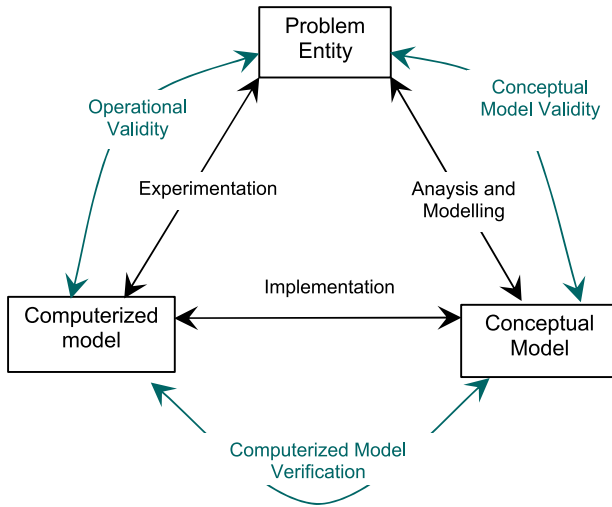


Fig. 1. Simplified version of the development process in technical simulation, according to Sargent (1998)

5 On the Relation of Verification and Validation to the Simulation Development Process

In the strict context of the simulation development process, could verification and validation acquire a less contentious meaning? Rigorously speaking, none of the terms can avoid acquiring significances dissociated from the philosophical problem of “truth”. In social science simulation, other terms are sometimes used to refer to the verification step, such as “internal validation”⁵ or “program validation”⁶. Notwithstanding, the usual terminology adopted from computer science and technical simulation seems to prevail, and is adopted by most researchers. References from simulation in computer science and engineering, such as Sargent (1998), are often used as the inspiration to define the terms. Whether the terminological mapping from technical to social science simulation is appropriate depends on how well the development process of technical simulation can be mapped to the development process of social science simulation. I will argue that this mapping can only be partial. However, I will argue that the usual terminology will remain.

In technical simulation, verification and validation are usually defined in terms of two kinds of models: the conceptual and the computerised model. That is, according to the simplified development process of Figure 1, a single conceptual model mediates between the problem entity and the computerized model. The problem entity is the ultimate subject of inquiry. The conceptual model is usually a mathematical/ logical/ verbal representation of the problem entity. Meanwhile, whereas the evaluation of technical simulations is essentially a quantitative analysis of the outputs, the evaluation

⁵ Axelrod (1997, p.27).

⁶ Richiardi et al. (2005, paragraph 4.28).

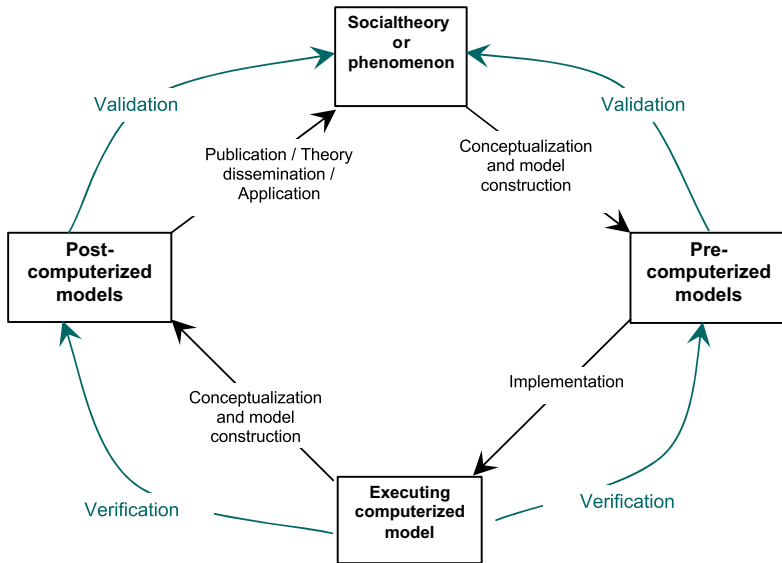


Fig. 2. The development process in social science simulation: representing explicit post-computerized models

of complex social models is also qualitative and highly conceptual. A fact that requires recognizing the construction of two distinct kinds of conceptual models during the simulation development process, one before the implementation and execution of the simulation programs – THE PRE-COMPUTARIZED MODEL – and another after the implementation and execution – THE POST-COMPUTARIZED MODEL. This requires recognizing that two subjects of inquiry exist in simulation, rather than one: the target theory or phenomenon and the executing computerized model, represented in Figure 2. In short, two conceptual models mediate between two subjects of inquiry.

The conceptual model on the right, designated here as the *pre-computerized model*, is a representation in the minds and writing of the researchers, which presumably represents the target social phenomenon or theory. This model must be implemented as a computerized executable model, by going through a number of intermediate models such as textual programs written in high-level programming languages. The inquiry of the executing model give rise to one or more conceptual models on the left, designated here as the *post-computerized models*. They are constructed based on the observation of execution behaviours and output data, often through graphing, visualisation and statistical packages. The whole construction process results in categories of description which may not have been used for describing the pre-computerized model. This gives rise to the usual idea of *emergence*, when interactions among objects specified through pre-computerized models at some level of description give rise to different categories of objects at different levels of description, observed in the executing model and specified accordingly through post-computerized models.

As a canonical illustration, consider the well known culture dissemination model of Axelrod (1997b) whose goal is to analyse the phenomena of social influence. At a certain level of description, the pre-computerized model defines the concept of *actors*

distributed on a grid, the *culture* of each actor is defined as a set of five features and *interaction mechanisms* are specified with a bit-flipping schema.⁷ The executing model gives rise to other categories of objects like the concepts of *regions* and *zones* on the grid, resulting from the interaction of several individual cultures, associated with properties of interest and conditions in which they form. A great deal of the simulation proposed by Axelrod concerns inquiring properties of *regions* and *zones* in the context of a new conceptual model proposed, such as the relation between the size of a *region* formed and the number of features per *individual culture*. These concepts are later interpreted in relation to the target social phenomena of social influence.

I will now situate the role of verification and validation in the simulation development sense. The goal is not to establish any kind of standard definition for these terms. Rather, the intention is to show that once the full set of different kinds of models are denoted explicitly in the development process, the mapping to the problem of “truth” in the philosophical sense is facilitated, insofar as the epistemological underpinnings of each model involved may be analysed – and thus run fewer risks of reducing epistemological debates to terminological ones. Recall that the reason for distinguishing the terms verification and validation in the development process is pragmatic, related to the need to determine the adequacy of several kinds of models against two distinct subjects of inquiry. In this context, consider the following definitions:

COMPUTERISED MODEL VERIFICATION is defined as checking the adequacy among conceptual models and computerized models. It is primarily concerned with ensuring that the pre-computerized model has been implemented *adequately* as an executable computerized model, according to the researcher and/or stakeholders’ intentions in the parameter range considered, and also that the post-computerized model *adequately* represents the executing model in the parameter range considered. You may now interject, What is the meaning of *adequate*? Indeed, this would return us to the epistemological debate, considered elsewhere, where adequateness can be explained according to a number of epistemological perspectives, such as the notions of formal, empirical and intentional knowledge.⁸ Our goal is essentially terminological. A minimal definition could be the following: adequateness means that the execution behaviours of the computerized model coincide with the semantics of both the pre- and the post-computerized models, in accord with the researcher’s intentions.

Finally, CONCEPTUAL MODEL VALIDATION is defined as ensuring that both the pre- and post-computerised models are *adequate* models of the target social theory or phenomenon. The term adequate, in this sense, may stand again for a number of epistemological perspectives, such as empirical or arbitrarily interpretative. Given the difficulty in establishing standards for the construction of social theory or the surveying of social facts, it is an epistemological issue beyond simulation with no easy solution. From a practical point of view, adequateness may be established according to the intuition of the researchers or the stakeholders involved in a participative-based simulation. In any case, the interest of validation is to assess whether the target social theory or phenomenon is adequately represented by relevant micro-levels of description of the pre-computerised model and relevant micro and macro-levels of description of the post-computerised model.

⁷ In which the probability of interaction between two actors is set proportional to a measure of similarity between two cultures.

⁸ See David et al. (2005). See also David et al. (2007).

Table 1. Some examples of common techniques and approaches for verifying or validating models in the simulation development process sense. Corresponding methodological perspectives are referenced as footnotes.

	Pre-Computerized Models	Post-Computerized Models
Validation	Formal Modelling Theory-driven Discourse Empirical Methods ⁹ Participative-based Approaches	Statistical Signatures Stylised Facts ¹⁰ Participative-based Approaches Cross-Model Validation (“Model-to-Model”) ¹¹
Verification	Structured Programming (e.g. modularity, object-oriented programming) Model Embedding ¹² (e.g. compilation, software reuse of components) Static Methods (e.g. code walk-throughs)	Dynamic Methods (program testing, sensitivity analysis) Replication for Alignment Participative-based Approaches

Table 2. A mapping between terminology and epistemological perspectives on simulation, related to validation and verification

Terminological and Epistemological Perspectives in Oreskes’ (1994) Sense	Epistemological Perspective in Isedas’ (1997) Sense	Subject of Inquiry	Terminological Perspective in the Simulation Development Process Sense	Conceptual Models Involved	Epistemological Perspectives in Terms of Kinds of Knowledge Involved
Can models be verified, validated or/and confirmed (in the broad philosophical sense)?	Can simulation yield knowledge?	Executing computerized model	Verification	Pre- and post-computerized models	Formal knowledge; Intentional and empirical knowledge (in David’s sense, 2005)
		Social theory or phenomenon	Validation		Multiparadigmatic

Given the simulation development process described, are there any crucial differences between verification point and validation in the simulation development process? From an epistemological point of view there are certainly differences, but from a terminological

⁹ See Boero and Squazzoni (2005).
¹⁰ See Bernd-O et al. (2005).
¹¹ See Hales et al. (2003).
¹² See David et al. (2005).

point of view it is a distinction that results from the computational methodology. Whereas *verification* is essentially a form of evaluating the logical inferences established between micro and macro concepts with reference to the computerised model, *validation* is essentially a form of evaluating the adequateness of *both* those logical inferences and concepts with close reference to the target social theory or phenomenon.

6 On Epistemological Perspectives

Can simulation yield knowledge about the social world? This is a philosophical question in the social sciences, but it is by no means the only important question. Whereas the body of research that simulation in the social sciences provides has been intense in the last two decades, the increase in philosophical literature came to the surface much more recently. Philosophical interest is an indication that simulation has become a consolidated discipline, with its influence on society, contributing with knowledge and critical thinking, with its own eclectic dilemmas, methods and techniques. Once we look into Figure 2, we may realize that a number of approaches and techniques can be identified for each quadrant of the simulation development process, described in Table 1. Details on these approaches and techniques are out of the scope of this essay, although some references on methodological and epistemological perspectives are indicated as footnotes. The point is that if simulation is to be analysed as a method, its established tools and techniques are to become ingredients of analysis, together with their role in the simulation development process. This will preclude us from running the risk of driving epistemological debates by means of terminological ambiguities.

What kind of knowledge can each model and technique provide? What is the range of purposes for these models and techniques, their methodological limits, ranges of application, what type of consensus can they achieve or provide? Once we realize that simulation is a complex type of model embedding¹³, and realize the meaning of verification and validation in the development process sense, the mapping from terminology and types of models to different categories of knowledge may become easier to investigate. Whether the question is whether simulation yields knowledge or the kind of knowledge that simulation provides, that is a philosophical debate that can be stated in the terms of Table 2, which illustrates a mapping between terminology and examples of epistemological perspectives, such as Oreskes et al. (1994), Iseda (1997) and David et al. (2005). An indication that acknowledges the consideration of explicit post-computerized models in the logic of the method of simulation, which helps us situate the roles of the terms verification and validation, and thus into the real epistemological debate.

References

1. Axelrod, R.: Advancing the Art of Simulation in the Social Sciences. In: Conte, R., Hegselmann, R., Terna, P. (eds.) *Simulating Social Phenomena*, pp. 21–40. Springer, Heidelberg (1997a)

¹³ In the sense of David et al. (2007).

2. Axelrod, R.: The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution* 41(2), 203–226 (1997b)
3. Heine, B.-O., Meyer, M., Strangfeld, O.: Stylised Facts and the Contribution of Simulation to the Economic Analysis of Budgeting. *Journal of Artificial Societies and Social Simulation* 8(4) (2005)
4. Boero, R., Squazzoni, F.: Does Empirical Embeddedness Matter? Methodological Issues on Agent Based Models for Analytical Social Science. *Journal of Artificial Societies and Social Simulation* 8(4) (2005)
5. David, N., Sichman, J., Coelho, H.: The Logic of the Method of Agent-Based Simulation in the Social Sciences: Empirical and Intentional Adequacy of Computer Programs. *Journal of Artificial Societies and Social Simulation (JASSS)* 8(4) (2005)
6. David, N., Sichman, J., Coelho, H.: Simulation as Formal and Generative Social Science: The Very Idea. In: Gershenson, C., Aerts, D., Edmonds, B. (eds.) *Worldviews, Science and Us: Philosophy and Complexity*, pp. 266–275. World Scientific, Singapore (2007)
7. Edmonds, B., Hales, D.: Replication, replication and replication: Some hard lessons from model alignment. *Journal of Artificial Societies and Social Simulation* 6(4) (2003)
8. Edmonds, B.: Towards an ideal social simulation language. In: Sichman, J.S., Bousquet, F., Davidsson, P. (eds.) *MABS 2002. LNCS*, vol. 2581, pp. 105–124. Springer, Heidelberg (2003)
9. Epstein, J.: Agent-Based Computational Models and Generative Social Science. *Complexity* 4(5), 41–59 (1999)
10. Fetzer, J.: The Role of Models in Computer Science. In: *General Topic: Philosophy of Computer Science*, La Salle. *The Monist*, vol. 82(1), pp. 20–36 (1999)
11. Gilbert, N., Troitzsch, K.: *Simulation for the Social Scientist*. Open University Press (1999)
12. Hales, D., Rouchier, J., Edmonds, B.: Model-to-Model Analysis. *Journal of Artificial Societies and Social Simulation* 6(4) (2003)
13. Iseda, T.: Verification, Validation, and Confirmation of Simulation Models: Suppe vs. In: Oreskes, et al. (eds.) *Assistantship at University of Maryland, PHIL 250 Philosophy of Science I* (instructor: Frederick Suppe), December 9 (1997) (retrieved March 2007), <http://carnap.umd.edu/phil250/oreskes/oreskes.html>
14. Küppers, G., Lenhard, J.: Validation of Simulation: Patterns in the Social and Natural Sciences. *Journal of Artificial Societies and Social Simulation* 8(4) (2005)
15. Oreskes, N., Shrader-Frechette, K., Belitz, K.: Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263, 641–646 (1994)
16. Richiardi, M., Leombruni, R., Sonnessa, M., Saam, N.: A Common Protocol for Agent-Based Social Simulation. *Journal of Artificial Societies and Social Simulation* 9(1) (2006)
17. Sargent, T.J.: Verification and validation of simulation models. In: *Proceedings of 1998 Winter Simulation Conference* (1998)

Validation and Verification of Agent-Based Models in the Social Sciences

Paul Ormerod and Bridget Rosewell

Volterra Consulting, London, UK
pormerod@volterra.co.uk

Abstract. This paper considers some of the difficulties in establishing verification and validation of agent based models. The fact that most ABMs are solved by simulation rather than analytically blurs the distinction between validation and verification. We suggest that a clear description of the phenomena to be explained by the model and testing for the simplest possible realistic agent rules of behaviour are key to the successful validation of ABMs and will provide the strongest base for enabling model comparison and acceptance. In particular, the empirical evidence that in general agents act intuitively rather than rationally is now strong. This implies that models which assign high levels of cognition to their agents require particularly strong justification if they are to be considered valid.

Keywords: Verification, validation, agent-based models, agents, behaviour.

“Where do correct ideas come from? Do they drop from the skies? No. Are they innate in the mind? No. They come from social practice, and from it alone; they come from three kinds of social practice, the struggle for production, the class struggle and scientific experiment.”

Mao Tse Tung, 1963

1 Introduction

This last kind of ‘social practice’ in the quote above, namely scientific experiment, is one of the few areas where the tradition of English empiricism blends with the more dogmatic nature of Marxist thought. In this paper, we argue that the best way to make progress in the validation of ABMs involves two clearly defined stages. The first is to construct correct criteria by which the output of a model is to be judged. The second is by exploration, by trying various approaches and seeing what works in practice to best meet these criteria. In this context, theory is led by practice. Abstract musing is of little or no value.

Agent based models (ABMs) in the social sciences are relatively new. In general they rely on numerical simulation to discover their properties instead of the more traditional route, within economics at least if not the other social sciences, of analytical solution. And numerical simulation in turn relies on the massively enhanced power of the personal computer. In short, ABMs are essentially a recent innovation which has only become feasible to implement over the past ten years or so.

For this reason, no firm conclusions have been reached on the appropriate way to verify or validate such models. We are in a period of rapid evolutionary change following an innovation. The innovation is ABMs, and the evolutionary process which is taking place involves a wide range of methodological issues around ABMs, one of which is the method of validation.

From the outset, it is important to emphasise that verification and validation are two distinct processes. Much of the literature focuses on validation, but verification is equally important. Pilch et al.(2000) and McNamara et al. (2007) characterise the distinction as follows:

- **Verification**
 - The process of determining that a computational software implementation correctly represents a model of a process
 - The process of determining that the equations are solved correctly
- **Validation**
 - The process of assessing the degree to which a computer model is an accurate representation of the real world from the perspective of the models intended applications
 - The process of determining that we are using the correct equations.

2 Verification

2.1 Simulation and Solution

We start with the issue of verification. This is essentially the question: does the model do what we think it is supposed to do? This is conceptually quite distinct from the issue of validation. Whenever a model has an analytical solution, a condition which embraces almost all conventional economic theory, verification is a matter of checking the mathematics.

But there is no inherent superiority in this approach to that of numerical simulation. An everyday example, in a scientific context, is the solution of partial differential equations by numerical methods. A notorious example of this is the proof of the Four Colour Problem, which asked whether a map could be coloured with only four colours so that no two adjacent areas had the same colour. No one has yet found an analytical solution but the advent of computers meant that a brute force simulation based solution could be produced by looking at over 2000 possible cases. Such a proof, written out, would require so long to check that it could not effectively be verified and not all mathematicians accept that this has the status of a proof at all (see Stewart 1996).

For most of human history, the inability to conduct such simulation tests has meant that analytical solution has been the only possible approach to model verification. Suppose for a moment, however, that the Greeks had invented the computer and statistical theory rather than algebra. We might now have not an analytical but a statistical concept of proof. Imagine that since the time of Antiquity there had been a stream of developments in computing power. A ‘proof’ of Fermat’s Last Theorem might well

be given by calculating that the relevant equality was not satisfied by any power (beyond 2) up to and including ten to the million raised to the million, a number unimaginably larger than any number which might ever be used in practice.

Mainstream economics has limited itself by its inability to go beyond analytical results. This also constrains the modelling strategy itself as we show when considering the validation issues.

The downside of ABMs in the verification process is that it is not as straightforward as with analytical solutions. We do not even have widely recognised routines such as the Runge-Kutta approach to numerical solutions of differential equations.

2.2 Replication

A key aspect of verification is replicability. Rand and Wilensky (2006) discuss this in detail, in particular the three criteria developed by Axelrod (1997). These are, first, numerical identity, showing that the original and replicated model produce exactly the same results. This seems less relevant now than perhaps it was ten years ago, because the behavioural rules in many ABMs typically contain stochastic elements.

The second point, distributional equivalence, covers this point. The properties of the original and replicated model should be statistically indistinguishable from each other. Note that the replication needs to be carried out over the same number of solutions as the original, especially when distributions of outcomes are right-skewed so that solving the replicated model for, say, two orders of magnitude more times than the original might generate a really extreme outcome not encompassed in the original results.

This in turn leads to the separate question of the number of times a stochastic ABM needs to be solved in order to establish its properties. There is little discussion of this in the literature, and in practice, certainly with small ABMs, 500 or 1,000 solutions are often quoted. This is an area which deserves more attention.

Axelrod's third point is 'relational alignment' by which is meant if input variable x is increased in both models then if output variable y increases in the original model it should also increase in the replicated model. Stated like this, the criterion seems rather weak, and should be extended to be 'if input variable x is increased in both models by a given amount, the distribution observed in the changes in output variable y should be statistically indistinguishable.

There is a parallel in replication with the world of econometrics. The ability to perform regressions, except of the simplest kind, was severely limited until the 1970s by access to computer power. There was then a proliferation of articles on applied econometrics, which turned into a torrent during the 1980s with the development of the personal computer. Initially, published papers carried the results but no data. Even the descriptions of data sources were often cursory if not cryptic. Replicating someone else's results was a time consuming task, which often failed (the career of one of the present authors started as an applied econometrician). Gradually, a better code of practice in describing data sources evolved, followed by an increasing insistence by the leading journals that the actual data used in the regressions (including all the transformations used) be made available to other researchers, initially in the paper itself and now via the web.

So econometrics has seen a gradual evolution of better replication practice, enabled by successive waves of technology. Even so, we conjecture that only a tiny fraction of published results are ever replicated by third parties, and these will be ones that in general emerge as being considered important by the applied econometric community¹. A similar sort of process will presumably take place with ABMs. The *Journal of Artificial Societies and Social Simulation*, for example, requires that the code for published models is made accessible to other researchers. This is an excellent move.

But for the model builders themselves, apart from checks within the group that the code appears to be written as intended, verification is by no means straightforward. A variant of Axelrod's 'relational alignment' is of some use. Namely, if input variable x is increased, does the direction (and possibly size) of the change in output variable y accord with prior beliefs. This is only a partial check, because one of the properties of an ABM may be the emergence at the macro-level of a phenomenon which could not be deduced from the micro-behavioural rules.

This also suggests that testing ABMs may blur the boundary with which we started between verification and validation. If we test the ABM in question by changing the various input variables to test the range of possible output values that the model produces then we have to have some means of judging whether that range is consistent with the model operating correctly or not. In most cases, this can only be done by considering the plausibility of the outcome with reference to the input ranges that have been chosen. This in turn will generally be based on the range of reality that the model is attempting to explain.

This is one reason why a clear description of the problem to be modelled and how the output should be judged is essential to both verification and validation.

3 Validation

3.1 Describing the Problem

It is easy to overlook this crucial stage of validation. Modellers often plunge into the difficulty of setting up a set of rules and building a model. Yet the process of validation requires a clear view of what the model is attempting to explain and for what purpose. What are the key facts that the model needs to explain and how well must it do it?

Without some answer to this question, it becomes impossible to judge the validation process. Since no model outputs will ever completely describe reality, then we need to know both something about what parts of reality we are trying to get most grip on.

There are likely also to be trade-offs in this description. A model which is trying to capture the range of consumer preferences will need to explain the variety of products but may be unable to precise about product innovation and growth. A model focused on how cities concentrate will focus on a measure of clustering that may have a very simplified geography.

¹ Occasionally, of course, a very bad but influential article will attract a lot of replication. A notorious example is a paper by Benjamin and Kochin in the *Journal of Political Economy* in the early 1980s which claimed that unemployment in inter-war Britain was caused by high benefits paid to the unemployed. It attracted many replications and rebuttals.

An aspect of a problem description which is often overlooked and becomes even more central for ABMs than other modelling strategies is that of time. ABMs are typically solved in steps. But what is the equivalent in real time with the step in any given ABM?

Conventional economic theory rarely if ever faces this crucial question. An analytical solution is obtained, and an equilibrium solved for a certain set of parameters and exogenous variables values. A change is posited, and a new equilibrium solution calculated. The time taken between these equilibria is almost never considered as an issue. A brilliant exception to this was published as early as 1969 by Atkinson. He showed, *inter alia*, that the typical time scale of transition from one equilibrium growth path to another in the Solow model was over 100 years. But this article appears to have been exorcised from student reading lists; its implication that economies, even in a strictly neo-classical world, spend a long time out of equilibrium presumably being too disturbing.

Economists often regard the difficulty of translating steps in a model into real time as a weakness of ABMs. But in fact this difficulty is a great strength, in two ways. First, mapping a step into a real time equivalent can be a useful part of the model calibration process. For example, Ormerod (2002, 2004) has a simple ABM from which several key empirical features of the US business cycle emerge. In calculating the size of the growth rate of aggregate output in the model, which emerges from the decisions of firms, the actual range experienced by the US economy during the 20th century is a key way of fixing a step in the model to be equivalent to a year in real time. Second, and more important, it is a valuable aspect of model validation. A sensible rationale has to be provided, in many applications, for the real time equivalent of each step in the model and this is an important of establishing what phenomena are under investigation. In the example quoted above, the model aimed to explain the characteristics of the business cycle over years, so the ability to replicate annual time steps was a useful test.

Another aspect of problem description is to establish what kind of agents are under consideration. As Windrum et al. (2007) note, economists have reacted to the success of ABMs by extending their own framework to incorporate certain aspects of, for example, agent heterogeneity, bounded rationality and increasing returns. All theories, even those like quantum physics which have been subjected to extremely rigorous empirical testing, are approximations to reality. The question is always, how reasonable are the approximations to the features of any particular problem being considered. By restricting themselves in general to models with analytical solutions, economists restrict the set of areas where their assumptions might be reasonable approximations to reality. A way of thinking about this question is the table below.

Table 1. Behavioural Models of Agent Cognition

Type of theory	Ability of agents to gather information	Ability of agents to process information
Rational	Full	Maximise
Bounded rational	Partial	Maximise
Behavioural rational	Partial	Rule of thumb

In standard theory, agents are assumed to gather all relevant information. This immediately restricts the range of problems which can be usefully examined to ones where the dimension is relatively low. Bounded rationality potentially extends this dramatically, by allowing only partial (and asymmetric) information gathering. But with these types of models, the assumption that the decision processing rule is one of maximisation further restricts the closeness of the approximation of the model to reality. Almost the entire discipline of psychology, for example, suggests that agents in general do not behave in this way.

Any individual modelling strategy has to decide what kind of agents are to modelled. Are the agents going to be required to process large amounts of information or can the problem be addressed by simpler ones.

In general, we consider that simpler agents with simpler rules are to be preferred. The simpler the rule, the easier it becomes to test the model and discover its implications. Occam's razor should apply. The key aspect to validation is that the outcomes of the model explain the phenomenon. If the model explains the phenomenon under consideration better than previous models do, it becomes the current best explanation. This is the best we can expect to do.

3.2 Testing the Outcomes

The key distinguishing feature of ABMs is that the macro properties of the system under consideration emerge from the behavioural rules which are assigned to agents at the micro level. Conventional economists are often resistant to ABMs because of the proliferation of decision making rules which appears to be possible. As Vernon Smith pointed out in his Nobel lecture, "Within economics there is essentially only one model to be adapted to every application: optimization subject to constraints due to resource limitations, institutional rules and /or the behavior of others, as in Cournot-Nash equilibria". But Smith followed this immediately with the sentence: "The economic literature is not the best place to find new inspiration beyond these traditional technical methods of modeling". So modelers developing the ABM tradition can hardly expect to find understanding within mainstream economics.

Both the present authors, and the experience is widely shared amongst ABM modelers, have encountered from economists a view which can be summarized as follows: you have presented one set of behavioural rules to explain your chosen phenomenon, but there must be many such sets which do this, so how do you know yours are correct? Some economists even go on to imply that it is easy to construct successful ABMs, an opinion which merely reveals their ignorance of the difficulties involved. The fact that they do not appear to appreciate that the rules of behaviour incorporated into much standard economic analysis are of a very special kind which does not stand up well to experiment should not blind us to the need to provide a sound basis for the choice of the rules of behaviour for any individual ABM application.

One key test is that the behavioural rules should be capable of justification using evidence from outside the model. The better this evidence, the more credible the rules. If evidence is absent, then the appropriate response is to create simple agents whose rules of behaviour are easy to understand. For example, if the agents are firms then there is a large body of evidence that profit seeking behaviour will be observed, even if it is modified and overlaid by other motivations. On the other hand, if the

problem is to consider how cities grow, where there will be multifarious agents whose rules of behaviour would be hard to describe, it may be better to test growth patterns by choosing simple agents, such as a unit of geography and allow it just to choose its mix of activities.

Another test is to compare the success in meeting a description of the phenomenon under consideration of different models. At present, there is little competition between ABMs attempting to explain the same set of macro phenomena. We are still at the stage where only one ABM exists which accounts for any given set of macro features.

This does not, however, mean that such an ABM is not scientific. On the contrary, the inherent methodology of ABMs is far more scientific than that of conventional economics. We identify a set of empirical macro features, and plausible behavioural micro rules are designed from which the macro properties emerge. This is a much more scientific methodology than econometrics, for example, much of which is mere curve fitting and is not modeling in any real sense of the word. And outcomes of ABMs can be compared with the explanations given by more traditional models.

In an important sense, the current process of building ABMs is a discovery process, of discovering the types of behavioural rules for agents which appear to be consistent with phenomena we observe. Once we leave behind the comfort blanket of maximization, we need to find out what works. We return to this point below.

3.3 Comparing Model Outcomes

One area where competing agent based models exist is the business cycle. Three are described here. All are successful in that from simple behavioural rules, a range of complex macro features of reality emerge. But they differ in scope and scale.

The simplest is the model of Ormerod (2002). In this model, the only agents are firms. We know that *ex post* most of the fluctuations in output over the course of the cycle arise from the corporate sector, so this is a reasonable approximation to reality. The agents are heterogeneous, both in size and in how they take into account uncertainty. They are myopic, and look only one period ahead with limited information. Firms use very simple rules to decide their rate of growth of output and their sentiment about the future, their ‘animal spirits’ in Keynes’ phrase. From these simple behavioural rules emerge several key features of the business cycle:

- Positive correlations in output growth between individual firms over the cycle
- The autocorrelation properties of real US output growth in the 20th century
- The power spectrum of real US output growth in the 20th century
- The exponential distribution of the cumulative size of recessions

The model of Dosi *et. al.* (2006) contains considerably more features. In addition to firms, workers/consumers also appear in the model. There are two types of firm, one of which carries out research and development and produces machine tools, the other buys machine tools and produces the consumption good. Profits and wages are determined within the model. There is technical change and firm entry and exit. This model embraces more features of capitalist economies, but the behavioural rules within the model remain simple. Rather, the agents operate under imperfect information, are myopic and use rules of thumb behaviour rather than maximization.

From this model emerge considerably more empirical phenomena observed in capitalism, but the set of phenomena reported does not overlap with the previous, much smaller model. An important extension of the Dosi model from that of Ormerod is that the latter assumes (with strong empirical justification) that most of the fluctuations in output arise from firms rather than from the consumption of the personal sector. This stylized fact emerges from the rules of the Dosi model rather than being assumed. In addition, the Dosi model is consistent with the following stylized facts:

- Investment and consumption tend to be pro-cyclical and coincident variables
- Employment and unemployment are lagging variables, the former being pro-cyclical and the latter anti-cyclical
- Productivity dispersion across firms is large at any point in time, and is persistent over time
- Firm size distributions are considerably right skewed
- Firm growth rate distributions are not Gaussian and can be proxied by fat-tailed tent-shaped densities

These two models, then, appear at first sight to be diverse. But closer inspection shows that they have a lot in common. In each model, a key feature of its ability to replicate macro phenomena is the heterogeneity amongst individual agents. The 'representative agent' of neo-classical theory has to be discarded completely. Further, in each model decision makers are myopic, and operate under both uncertainty and imperfect information. It is as if decision makers in each of the models operates with low cognition. They are short-sighted, gather uncertain and limited information, and use simple rules with which to make decisions. The results of both models also contain a more complete description of the business cycle than non agent based models.

The model of Wright (2005) reduces the level of cognition assigned to agents even further. The model has nine rules which attempt to capture the social relations of production under capitalism. Almost all of these are purely stochastic, so it is as if agents are operating with zero cognitive ability. The deterministic aspect of the model is simply that firms must have enough money to pay the labour force, and if not they fire workers (at random) until this constraint is satisfied. In the model, a small class of capitalists employ a large class of workers organized within firms of various sizes that produce goods and services for sale in the marketplace. Capitalist owners of firms receive revenue and workers receive a proportion of the revenue as wages.

Again, the stylized facts which emerge from the model contain some overlap with the Dosi *et. al.* (2006) model, but additional ones as well:

- The power law distribution of firm size
- The Laplace distribution of firm growth
- The lognormal distribution (when aggregated over time) of firm extinctions
- The lognormal-Pareto distribution of income
- The gamma-like distribution of the rate of profit of firms

Deciding which of these models is more valid (we assume that they are equally verified) is not straightforward. They each produce results which are consistent with what is observed and do so with agent decision rules which are easy to describe and explain. The elements of reality which are explained are in each case more comprehensive than the standard model but add different levels of explanation.

3.4 Simple Agents, Simple Rules

One criterion that might be used to determine performance is simplicity of behaviour, on the principle that if simple agent rules can produce a good description, this is better than having complicated ones.

Another way of expressing this is to ensure that agents are only required to have the minimum necessary ability to process information or to learn. The issue of low, or even zero, cognition of agents is, we believe, an important aspect of the validation of ABMs. The idea that agents act, in general, in a way far removed from the precepts of rational choice economics is now supported by an impressive body of evidence (i.e. Kahneman 2003). It is models which assign, implicitly or explicitly, a high level of cognition to agents which need special justification rather than those which do not.

We suggest that a model that can only work if agents have high levels of cognition (information processing and learning) will only be able to capture a limited set of outcomes. Indeed one way of testing an ABM in the social sciences is to assign increasing levels of cognition to agents to see at what point the model ceases to provide a description of reality.

An example here is Ormerod and Roswell (2003). Two key stylised facts have been established about the extinction patterns of firms. First, the probability of extinction is highest at the start of the firm's existence, but soon becomes more or less invariant to the age of the firm. Second, a recent finding, that the relationship between the size and frequency of firm extinctions is closely approximated by a power law. The model is in the spirit of the Wright model, in that the focus is on the structural relationships between firms rather than on specific rules of agent behaviour.

In the basic model, firms by definition have zero cognition. Firms are assigned at random an initial level of fitness. A matrix, whose elements are initially chosen at random, describes both the sign and the size of the impact of all other firms on the fitness of any given firm. The actions of a firm can increase the fitness of another (e.g. if it is an efficient supplier to that firm) or decrease it (e.g. if they are direct competitors in the labour market). Despite the stark simplicity of the model, it contains a key feature of reality which is absent from most of conventional economics. The focus of the latter is upon competition between firms, whether in the output or the labour markets. But in reality, a great deal of economic activity is business-to-business in which the success of any given firm is beneficial to those which it supplies.

In each step of the model, one of the elements of the matrix of connections is updated at random for each firm. A firm is deemed extinct if its fitness falls below zero. So the firms are unable by definition to acquire knowledge about either the true impact of other firms' strategies on its own fitness, or the true impact of changes to its own strategies on its fitness.

From this model with simple rules of behaviour, the observed stylised facts emerge. We then go on to examine the effects of allowing firms different amounts of knowledge about the effects of strategy in the context of the agent-based evolutionary model. There are very considerable returns in the model to acquiring knowledge. There is a sharp increase in the mean agent age at extinction for agents with even a small amount of knowledge compared to those without. Indeed, we find that as both the amount of knowledge available to firms increases and as the number of firms capable of acquiring such knowledge rises, the lifespan of agents begins to approach

the limiting, full information paradigm of neo-classical theory in which agents live for ever. However, even with relatively low levels of knowledge and numbers of agents capable of acquiring it, the model ceases to have properties which are compatible with the two key stylised facts on firm extinctions. The clear implication is that firms have very limited capacities to acquire knowledge about the true impact of their strategies.

By starting with simple rules, we establish that key facts emerge from the interactions of the model. Giving the agents more complicated rules in fact turns out to destroy this feature. Only by starting with the simplest possible configuration could this be discovered.

4 Conclusion

Agent Based Models face a variety of issues in verification and validation which are new, precisely because ABMs offer the opportunity to model a wider class of phenomena than has been possible before.

Because such models are based on simulation, the lack of an analytical solution (in general) means that verification is harder, since there is no single result the model must match. Moreover, testing the range of model outcomes provides a test only in respect to a prior judgment on the plausibility of the potential range of outcomes.

In this sense, verification blends into validation. We take the Popperian position that validation can never be proof and that we are therefore seeking for models which explain more than their predecessors and are not falsified. Thus we stress that an important part of validation is a clear description of what is being explained. This should also ideally include a description of what is not explained by the current best practice.

Our second stress is on simplicity. The validation of models with complicated agents and complicated rules is impossible in our current state of knowledge. It may be that over time the validation of simpler models will then lead to the ability to make them more complicated as levels of validation are built up. Proofs in mathematics now often require this nested approach.

However, agent based models have not yet reached this position and in our view will not do so unless there is rigorous testing of the simplest possible models at the outset. Thus we do not believe that a model with complicated agents should be accepted unless and until it has been shown that simpler (lower information, lower cognition, less processing, less learning) ones will not explain the phenomenon just as well. In particular, we believe that the growing empirical body of evidence which shows that in general agents act in an intuitive rather than rational way is an essential part of model validation. Models which assign high levels of cognition to their agents need particularly strong justification if they are to be considered valid.

References

1. Atkinson, A.B.: The Timescale of Economic Models: How Long is the Long Run? *Review of Economic Studies* 36, 137–152 (1969)
2. Axelrod, R.: Advancing the Art of Simulation in the Social Sciences. In: Conte, R., Hegselmann, R., Terna, P. (eds.) *Simulating Social Phenomena*, Berlin, pp. 21–40. Springer, Heidelberg (1997)

3. Dosi, G., Fagiolo, G., Roventini, A.: An Evolutionary Model of Endogenous Business Cycles. *Computational Economics* 27(1), 3–34 (2006)
4. Kahenman, D.: Maps of Bounded Rationality: A Perspective on Intuitive Judgement and Choice. *American Economic Review* 93, 1449–1475 (2003)
5. McNamara, L., Trucano, T., Backus, G.: Verification and Validation as Applied Epistemology. In: Proceedings of the UCLA Conference on Human Complex Systems (April 2007), <http://hcs.ucla.edu/arrowhead.htm>
6. Ormerod, P.: The US business cycle: power law scaling for interacting units with complex internal structure. *Physica A* 314, 774–785 (2002)
7. Ormerod, P.: Information cascades and the distribution of economic recessions in capitalist economies. *Physica A* 341, 556–568 (2004)
8. Ormerod, P., Rosewell, B.: What can firms know?, plenary talk. In: Proc. NAACSOS conference, Pittsburgh (2003)
9. Ormerod, P., Rosewell, B.: On the Methodology of Assessing Agent Based Models in the Social Sciences. In: Metcalfe, J.S., Foster, J. (eds.) *Evolution and Economic Complexity*, Edward Elgar, Cheltenham (2004)
10. Pilch, M., Trucano, T., Moya, J., Groehlich, G., Hodges, A., Peercy, D.: Guidelines for Sandia ASCI Verification and Validation Plans – Content and Format: Version 2.0., Albuquerque, NM: Sandia National Laboratories, SAND 2000–3101 (2000)
11. Rand, W., Wilensky, U.: Verification and Validation through Replication: A Case Study Using Axelrod and Hammond’s Ethnocentrism Model. In: Proc. NAACSOS (2006)
12. Rosewell, B., Horton, A.: Neighbours are not enough. In: *European Conference on Complex Systems*, Oxford (2006)
13. Smith, V.L.: Constructivist and Ecological Rationality in Economics. *American Economic Review* 93, 465–508 (2003)
14. Stewart, I.: *From Here to Infinity*. Oxford University Press, Oxford (1996)
15. Windrum, P., Fagiolo, G., Moneta, A.: Empirical Validation of Agent-Based Models: Alternatives and Prospects. *Journal of Artificial Societies and Social Simulation* 10, 2 (2007)
16. Wright, I.: The social architecture of capitalism. *Physica A* 346, 589–620 (2005)

Abductive Fallacies with Agent-Based Modeling and System Dynamics

Tobias Lorenz

Stiftung Wertevolle Zukunft, Alte Rabenstraße 32, 20148 Hamburg
tobias.lorenz@gmail.com

Abstract. Increasing usage of computer simulation as a method of pursuing science makes methodological reflection immanently important. After discussing relevant philosophical positions Winsberg's view of simulation modeling is adapted to conceptualize simulation modeling as an abductive way of doing science. It is proposed that two main presuppositions determine the outcome of a simulation: theory and methodology. The main focus of the paper is on the analysis of the role of simulation methodologies in simulation modeling. The fallacy of applying an inadequate simulation methodology to a given simulation task is dubbed 'abductive fallacy'. In order to facilitate a superior choice of simulation methodology three respects are proposed to compare System Dynamics and Agent-based Modeling: structure, behavior and emergence. These respects are analyzed on the level of the methodology itself and verified in case studies of the WORLD3-model and the Sugarscape model.

Keywords: Abduction, System Dynamics, Agent-based Modeling, Methodology, Multi-Paradigm Modeling.

1 Methodological Preliminaries

Working with computer simulations as a method of scientific research is becoming more and more common in the social sciences and gives rise to new fields such as experimental pragmatics, socionics or Agent-Based computational economics. This makes methodological reflection upon these new disciplines immanently important.

In order to understand the philosophy of science of computer simulation some basic concepts have to be clarified. A simulation is to be understood here as a parameterized instantiation of a model, i.e. a generated time series. A model comprises more. It is to be discriminated from the program code because it is more abstract and temporally prior to it [22]. It constitutes the totality of possible simulations. This model as well as the simulation rests meaningless without some semantic framework around it. The model itself is assumed to be influenced by two elements: theoretical foundations and computational methodology. The theoretical foundations are composed by causal relationships which are formulated in a scientific field. Methodology stands for the chosen way to translate that theory into a simulation model. While being translated into a simulation model the theory is adapted to the chosen methodology. Simulation modeling is to mean the whole process of doing research via simulations. Nevertheless the methodology of this simulation modeling process remains

unclear and has to be assessed: “These methods are called ‘simulations’, or ‘numerical experiments’; names that strongly evoke the metaphor of experimentation. At the same time, the mathematical models that drive these particular kinds of simulation are motivated by theory. Prima facie, they are nothing but applications of scientific theories to systems under the theories’ domain. So where ‘on the methodological map’ do techniques of computer simulation lie?” [24]

A recent discussion within philosophy of science tries to answer this question. It is part of the broader discussion about the methodological role of models in general – not only computational models intended for simulation – which can neither be easily classified with the existing categories of philosophy of science. One prominent position held by Mary Morgan and Margaret Morrison sees models as “mediating instruments” [13] or “autonomous agents” [14]. Thereby they want to make explicit that building models is a fruitful way of pursuing science which is neither purely theory-driven nor purely data-driven. This partial independence of both world and theory allows models to mediate between theory and data.

In another article Morgan analyses the effect of an increased fraction of immateriality in experiments through computer support. Apart from traditional laboratory experiment and purely mathematical computation she discriminates between “‘virtually’ experiments, ones in which we have nonmaterial experiments on (or with) semi-material objects, and ‘virtual’ experiments, ones in which we have nonmaterial experiments but which may involve some kind of mimicking of material objects” [12]. She concludes:

“The qualities of these in-between hybrids turn out to run along several dimensions. They involve a mix of models and experiments. They mix mathematical and experimental modes of demonstration. They mix types of control by assumption with experimental controls. They mix nonmaterial and material elements. They represent the world via mixed modes of representation and representativeness. [...] I have taken materiality to be one of the prime characteristics of interest in these hybrids and suggest that when faced with such a hybrid experiment, we need to look carefully at where and how much materiality is involved, and where it is located, before we can say much more about its validity.” [12]

Though Morgan does not give a systematic definition of materiality¹, the opposition between material and immaterial experiments rests intuitively clear. Immaterial experiments bring us closer to the phenomenon to be analyzed here, computer simulation. Simulation modeling shares this hybrid character of Morgan’s hybrid experiments.

According to Eric Winsberg simulation modeling is neither mere calculation of theoretically motivated equations nor real experimentation. It departs from mere calculation of theoretically motivated equations by its use of extra-theoretical techniques to calculate (e.g. discretization) and to draw conclusions from the simulation. It departs from real experimentation “because it assigns experimental qualities only to

¹ Morgan tries to get hold of the notion of materiality through the following passage: “From the use of a vacuum pump to supercolliders, however artificial the environment that is created, however artificial the outcome, the experimental intervention itself involves an action upon or the creation of a material object or phenomenon. In contrast, modern economics tends, in the main, to function by using extended thought experiments in which the material domain remains absent.” [12].

those aspects of simulation reasoning that occur *after* it is assumed that the simulation algorithm ‘realizes’ the system of interest.” [24] Simulation modeling diverges from both ways of pursuing science and is independent because there is a tradition of techniques of how to carry out simulations, which have a ‘life of its own’. These techniques of simulation modeling comprise “the whole host of activities, practices, and assumptions that go into carrying out a simulation.” [24]

As these techniques are being partly based on subjective experience and partly on customs of a community, simulation modeling is more creative but also more local than traditional ways of doing science. It is due to this fact that simulation results are only accepted relative to a scientific community and methods for validation are hard to define [16]. In addition these techniques are usually incorporated on a pragmatic basis to produce simulations.²

Thereafter these techniques “carry with them their own history of prior successes and accomplishments, and, when properly used, they can bring to the table independent warrant for belief in the models they are used to build.” [24]

After these techniques have made their way into the tradition of a community they have a tendency to be pursued unreflectedly, contrary to calls to reflection.³ If techniques are being accepted without methodological reflection upon their adequacy, they become dogmatic and constitute a paradigm. A paradigm is here to mean a set of assumptions which is to a large extent not questioned within its scientific community. The point to be stressed here is that these paradigms are relative to a scientific community. Donella Meadows and Jennifer Robinson for example postulate that “Different modeling paradigms cause their practitioner to define different problems, follow different procedures, and use different criteria to evaluate the results.” [11]

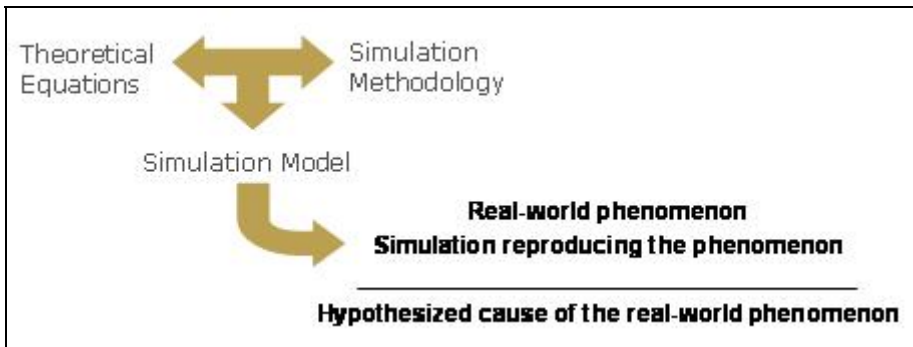


Fig. 1. Idealized version of the modeling process

² Compare the following for a general pragmatic account of modeling: “On this way of thinking, the scientific practices of representing the world are fundamentally pragmatic. [...] S uses X to represent W for purposes P. Here S can be an individual scientist, a scientific group, or a larger scientific community. W is an aspect of the real world. So, more informally, the relationship to be investigated has the form: Scientists use X to represent some aspect of the world for specific purposes”. [5]

³ “And we argue against these: [...] that any particular technique (including agent-based simulation) will always be appropriate for all modeling tasks, rather the domain should guide the choice of technique from a large palette of possibilities”. [15]

In order to understand better the effect of these paradigms in the process of computer simulation an idealized description of this process is to be developed. To simulate is to parameterize a model in order to reproduce data of a real world phenomenon by a simulation. This means to generate a certain effect representing a real phenomenon within a simulation. Having achieved this, the simulation is used to abduce the cause of the real world phenomenon.

Simulation modeling is an abductive process. Also classical science runs the risk of applying theories inadequately to a phenomenon. What is to be focused on here is the additional risk of applying an inadequate simulation methodology for translating theoretical statements into a computable model⁴. The choice of an inadequate methodology causes an abductive fallacy.

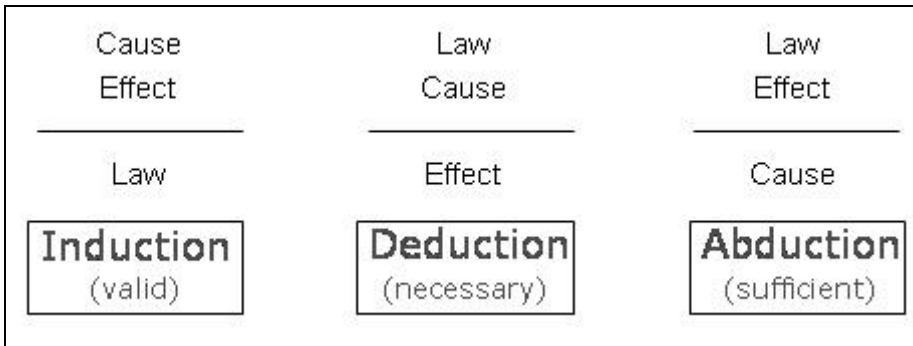


Fig. 2. Induction, deduction and abduction

Induction infers a valid law through the repeated observation of cause and effect. Deduction infers a necessary effect from a law and an observed fact. Abduction on the other hand infers a hypothesized cause for an explanandum via a simulation. This way of doing science can only make an explanation plausible. Due to the fact that abductive inferences can infer sufficient causes but not necessary ones [4], there is a risk to apply an inadequate simulation for explanation. This application of a wrong simulation to a given real-world phenomenon constitutes a fallacy. If the model underlying the simulation is derived from a problem-adequate theory but an inadequate methodology, it is an abductive fallacy in the sense employed here. To get a clearer perception of possible points of fallacy the constitutive elements of a model have to be developed. Winsberg proposes a list of elements, which need to be taken into account in order to transform theoretical structures into a computer simulation model:

- A calculational structure for the theory.
- Techniques of mathematical transformation.
- A choice of parameters, initial conditions, and boundary conditions.

⁴ “Successful numerical methods, therefore, invariably require of the simulationists that they transform the model suggested by theory in significant ways. Idealizations, approximations, and even self-conscious falsifications are introduced into the model. In the end, the model that is used to run the simulation is an offspring of the theory, but it is a mongrel offspring. It is also substantially shaped by the exigencies of practical computational limitations and by information from a wide range of other sources”. [24]

- Reduction of degrees of freedom.
- Ad hoc models.
- A computer and a computer algorithm.
- A graphics system.
- An interpretation of numerical and graphical output coupled with an assessment of their reliability. [23]

In the following two different methodologies which have led to separate communities rarely taking note of each other shall be compared. Literature regarding criteria for choosing between these two methodologies is hard to find [1], [9]. The point of departure between these communities depends mostly upon the way to reproduce a social phenomenon in terms of models, it depends upon a calculational structure for the theory and techniques of mathematical transformation.

2 System Dynamics and Agent-Based Modeling⁵

Two methodologies shall be analyzed with respect to their influence upon the development of simulation models: System Dynamics and the younger field of Agent-Based Modeling.

System Dynamics is based on differential equations and tries to capture systems with the so-called stock-and-flow-notation. This notation singles out aggregations and analyzes their change through feedback mechanisms. In Agent-Based Modeling, “the individual members of a population such as firms in an economy or people in a social group are represented explicitly rather than as a single aggregate entity.” [21]. “This massively parallel and local interactions can give rise to path dependencies, dynamic returns and their interaction.” [7]

Through its focus on individual entities, Agent-based approaches can be characterized as follows. They are suitable to:

- a) describe and demonstrate how the interaction of independent agents create collective phenomena;
- b) identify single agents whose behavior has a predominant influence on the generated behavior;
- c) identify crucial points in time, at which qualitative changes occur. [7]

Both System Dynamics and Agent-based Modeling are regularly utilized to explain socio-economical phenomena but differ in the way they approach their explanandum. System Dynamics typically looks for a reference mode of the central variable (which is to be reproduced and explained), whereas in Agent-based models an agent with individual behavior is being modeled and the emergent behavior arising out of the interaction of a population of those agents is being observed. Through this fact one might discriminate System Theory from Complexity Theory as being either confirmatory or exploratory [18]. Nevertheless both methodologies can be characterized as abductive, since their intention is to find explanations for given phenomena via simulation.

Nadine Schieritz and Peter Milling have compared Agent-based Modeling and System Dynamics by the following criteria:

⁵ For a more comprehensive analysis also including Discrete Event Simulation compare [9].

Table 1. Comparison of System Dynamics and Agent-Based Modeling [20]

	System Dynamics	Agent-based Simulation
Basic building block	Feedback loop	Agent
Unit of analysis	Structure	Rules
Level of modeling	Macro	Micro
Perspective	Top-down	Bottom-up
Adaptation	Change of dominant structure	Change of structure
Handling of time	Continuous	Discrete
Mathematical formulation	Integral equations	Logic
Origin of dynamics	Levels	Events

Although this list is excellent to inspire reflection it does not offer a satisfying concept of discrimination of both methodologies.⁶ Other directions to discriminate both methodologies can be found in the diverging approaches to individuals and observables [17] or the concept of emergence. Here the methodologies shall be discriminated with respect to three dimensions: structure (How is the model built?), behavior (What are the central generators of behavior?) and emergence⁷ (Can the model capture emergence?).

System Dynamics and Agent-based Modelling diverge in a number of points, which also put the notion of “structure” of a model to discussion. The **structure** of a model built according to the System-Dynamics-Methodology is static, whereas in an Agent-Based model, structure is dynamic, i.e. it changes over time. This is constituted by the fact that in Object-oriented programming new objects can be instantiated while running the simulation thereby changing the structure of the model. In System Dynamics the structure is being developed in advance and then used for mere calculation of the parameters. In addition System Dynamics and Agent-based Modeling differ in the number of levels they model. Whereas Agent-based Modeling comprises at least a micro level and a macro level, System Dynamics only models on one level, it stays ‘flat’.⁸

Furthermore both methodologies diverge in the elements which are supposed to generate **behavior**⁹. Two assumptions about the elements generating behavior are regarded as central in System Dynamics: a) feedback is central in generating behavior and b) accumulations are central in generating behavior.

Agent-based Modeling relies on a different set of basic assumptions: a) Micro-Macro-Micro feedback is central in generating behavior and b) interaction of the systems elements is central in generating behavior.

⁶ For a detailed critique of the Schieritz/ Milling approach compare [8].

⁷ Emergence here refers to the definition by Jeffrey Goldstein as interpreted by Peter Corning, who cites Goldstein’s concept of emergence as “the arising of novel and coherent structures, patterns and properties during the process of self-organization in complex systems.” Emergence is characterized by (1) radical novelty, (2) coherence or correlation, (3) a global or macro “level”, (4) evolution, (5) perceivability; [3].

⁸ Compare “[...] the ‘number of levels’ refers to whether the techniques can model not just one level (the individual or the society), but the interaction between levels.” [6], System Dynamics is characterized as having only one level.

⁹ For a detailed discussion compare [9].

Regarding a) one sees that both methodologies somehow incorporate feedback. But the feedback differs: In System Dynamics – due to the fact that it incorporates only one level of modeling – the feedback is ‘flat’ whereas in Agent-based Modeling – incorporating at least two levels – there is inter-level feedback. R. Keith Sawyer works this out as emergence and downward causation [19].

This leads to another point of departure between the methodologies: the concept of **emergence**. Emergence is made possible by the multilevel structure of Agent-based Modeling whereas the mono-level structure of System Dynamics is insufficient for emergence. Whereas the latter observes the same level it also models, the former models one level - the micro level - and analyzes another level - the macro level. The phenomena emerging on the latter can then be related to the algorithms of the micro level. The causal relationship between cause (on the micro level) and effect (on the macro level) seems less tight compared to the causal relationship between variables on the same level as in System Dynamics. As emergence presupposes at least two ‘levels’ (micro and macro) and System Dynamics as a methodology only works with one level, emergence is not possible in System Dynamics models.¹⁰

3 Structural Case-Study

These conceptual differences are now to be exemplified in the structure of two simulation models. Two outstanding examples of both fields have been chosen: the Sugarscape model as described in “Growing artificial societies” [4] and the WORLD3-model as described in “Limits to growth” [10] for the analysis of a System Dynamics model.

The Sugarscape model in its basic form consists of a square with a length of 50 spots. Each spot has a level of sugar and a sugar capacity and can host one agent. Sugar is being harvested by the agents and replaced by new sugar every time step. Each agent has two central properties: its metabolism and its vision. The metabolism is the amount of sugar the agent needs per time step and the vision is the number of spots horizontally as well as vertically which the agent can perceive. The agents hold an initial supply of sugar and can stock up without limits. Movement of the agents is regulated by a set of movement rules, which is iterated every time step.

As developed above the multilevel **structure** is a prominent property of Agent-based models. It is now central to see that whereas rules are developed for the individual, the focus of the observation is on a different level, it is on the macro-structure emerging out of these individual rules. As the authors themselves state: “Understanding how simple local rules give rise to collective structure is a central goal of the sciences of complexity” [4].

Chapter III of their book introduces sexual reproduction of the agents¹¹ into the model and through the instantiation of new agents a more dynamic structure. Every new agent is being created as a new object. This new element of evolution “gives rise to a rich variety of global, or macroscopic, population dynamics” [4].

¹⁰ “A technique capable of modelling two or more levels is required to investigate emergent phenomena.” [6]

¹¹ Compare Agent sex rule S, it states among others “If the neighbour is fertile and of the opposite sex and at least one of the agents has an empty neighboring site (for the baby), then a child is born”; [4].

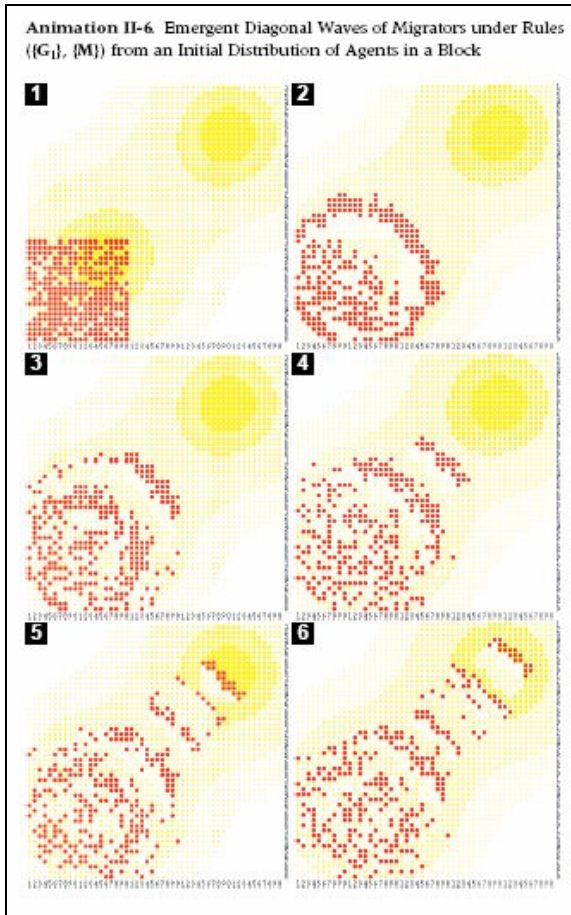


Fig. 3. Typical Animation of the Sugarscape Model [4]

“Indeed, the defining feature of an artificial society model is precisely that fundamental social structures and group behaviors emerge from the interaction of individual agents operating on artificial environments under rules that place only bounded demands on each agent’s information and computational capacity” [4]. The feedback from the macro-structure back down to the micro-level is a little harder to grasp, but as Joshua Epstein and Robert Axtell state in a footnote: “The term “bottom up” can be somewhat misleading in that it suggests unidirectionality: everything that emerges is outside the agent. But in models with feedback from institutions to individuals there is emergence inside the agents as well” [4]. The theses regarding **behavior** and **emergence** are therefore justified for this specific model. Picture 3 is a graphical representation of the macro level showing the effect of some specific micro level algorithm.

The WORLD3-model utilized in “Limits to growth” is based on the System Dynamics. It interconnects de- and increasing aggregations via feedback loops. These interconnections are based on causal relationships. As there are plenty of such relationships in the WORLD3-model it is harder to grasp in its completeness. In order to get an impression the demography sector is shown below.

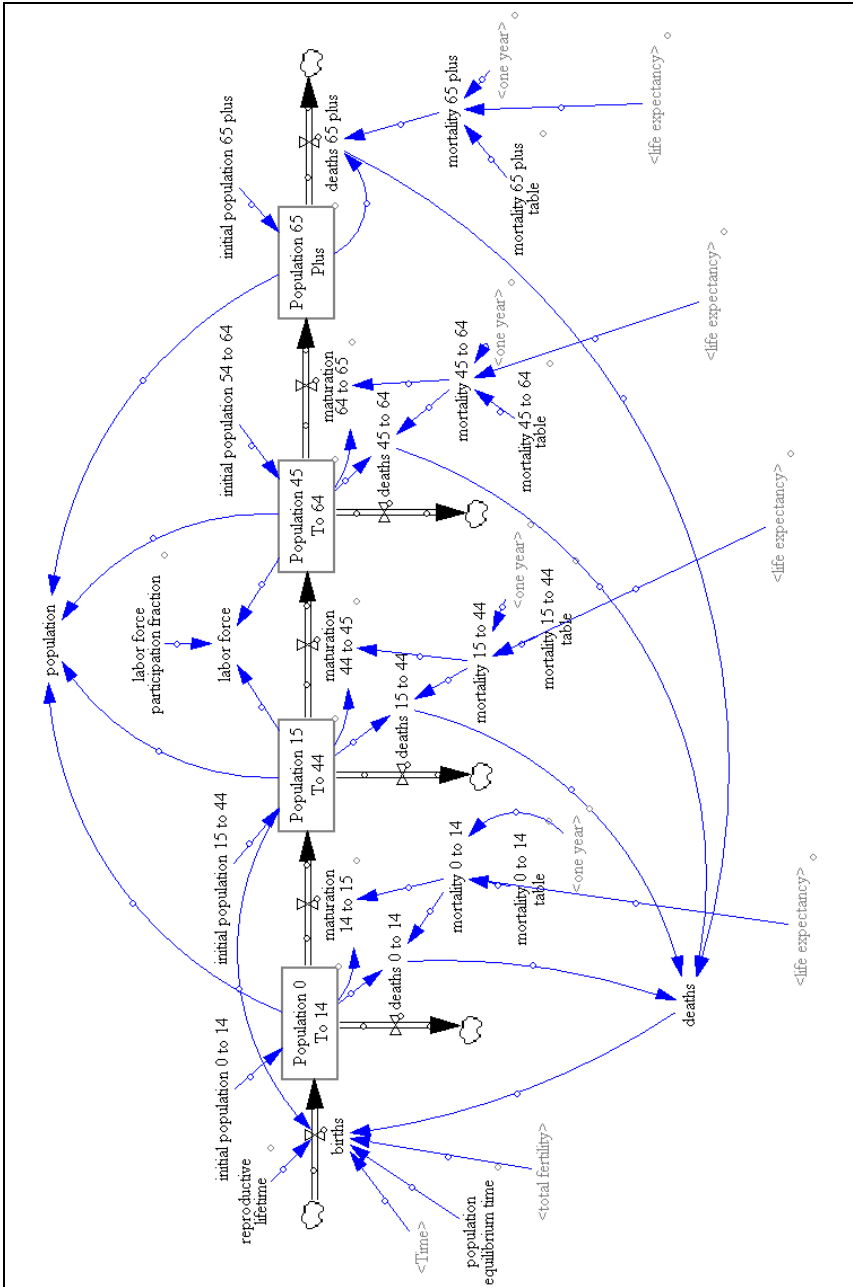


Fig. 4. Demographics Sector of the World3 Model¹²

¹² The WORLD3-version coming along with the VENSIM® library has been utilized here.

This sector consists of an aging chain, modeling different age cohorts and causal relationships, e.g. life expectancy, influencing the in- and outflows of the cohorts. What is being observed in the System Dynamics model is the change over time in the behavior of the modeled variables themselves.

So whereas in Agent-based modeling at least two levels are necessary (micro and macro), in System Dynamics there is only one level, whose dynamics are observed. In this example, the behavior of the variable ‘population’ which is shown in the Stock-and-Flow-diagram above is also subject to observation and analysis. The **structure** is ‘flat’.

The **behavior** of the WORLD3-model is based on the feedback thought. “We can begin our dynamic analysis of the long-term world situation by looking for the positive feedback loops underlying the exponential growth in the five physical quantities we have already mentioned” [10]. Apart from positive feedback negative feedback is also incorporated. As developed above, due to the ‘flat’ structure of System Dynamics models, **emergence** is not incorporated.

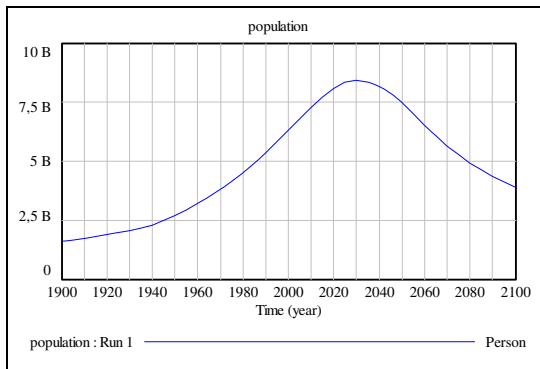


Fig. 5. Output of a System Dynamics Model

4 Beyond Fallacy

One area, in which both methodologies seem suitable, is diffusion dynamics [4], [21]. Whereas System Dynamics stresses the feedback aspect of diffusion, Agent-based Modeling stresses the interactional aspect of diffusion. In addition Agent-based Modeling has the advantage of being able to represent spatially. Nevertheless the question about the adequate simulation methodology remains open. Even a guiding framework for the characterization of the adequate methodology is out of hand. On the other hand the choice of methodology subjectively by the individual modeler seems insufficient.

In order to make simulation modeling more objective, the premises of the existing methodologies have to be worked out in detail to derive a set of situational characteristics which define the adequacy of a specific methodology in a given situation.

In addition to the risk of an abductive fallacy, the advance of graphical interfaces gives rise to another fallacy, which is constituted by the misinterpretation of data generated by a simulation model as empirical data. This fallacy can be analyzed following Hans Vaihinger’s theory of fiction and shall be dubbed realistic fallacy. The

more realistic graphical output of simulation modeling becomes, the harder it gets to grasp its being generated by a simulation model. “So psychologically, at the very least, working with a simulation is much more like doing an experiment if the simulation produces life-like images reminiscent of laboratory photographs” [24].

Rigorous care about the underlying sources of such data is necessary in order to avoid this fallacy.

5 Conclusion

In order to overcome abductive fallacies rigorous methodological reflection is necessary while working with computer simulation. Therefore criteria for when to apply given methodologies are necessary. One step towards a framework for these criteria might be to discriminate between the phenomenon (What is being modeled?) and the purpose (Why is it being modeled?) of a model. The characteristics of the methodology have to fit these two. Therefore these questions have to be answered rigorously *before* choosing the methodology to be used for the simulation model. According to the answers the methodology can be selected through its inherent characteristics: Agent-based Modeling might then be suitable for phenomena, which are governed by interacting entities, the necessity of a spatial distribution and heterogeneity of the individuals, whereas System Dynamics would be suitable for phenomena governed by ‘flat’ feedback and nonlinearities [9].

References

1. Brassel, K., Möhring, M., Schumacher, E., Troitzsch, K.: Can agents cover all the world? In: Conte, R., Hegselmann, R., Terna, P. (eds.) *Simulating social phenomena*, pp. 55–72. Springer, Heidelberg (1997)
2. Casti, J.: *Would-Be Worlds: How simulation is changing the frontiers of science*. John Wiley & Sons, New York (1997)
3. Coming, P.: The re-emergence of ‘emergence: a venerable concept in search of a theory. *Complexity* 7(6), 18–30 (2002)
4. Epstein, J., Axtell, R.: *Growing artificial societies*. MIT Press, Washington (1996)
5. Giere, R.: How models are used to represent reality. *Philosophy of science* 71, 742–752 (2004)
6. Gilbert, N., Troitzsch, K.: *Simulation for the social scientist*. Open University Press, Buckingham (1999)
7. Gebel, T., Pyka, A.: Agent-based modelling – A methodology for the analysis of qualitative development processes, 2004. In: Lombardi, M., Squazzoni, F. (eds.) *Saggi di economia evolutiva*, FrancoAngeli, Milano (2005)
8. Lorenz, T., Bassi, A.: Comprehensibility as a discrimination criterion for Agent-Based Modelling and System Dynamics: An empirical approach. In: *Proceedings of the 23rd International Conference of the System Dynamics Society, Boston (2005)*, <http://www.systemdynamics.org/conferences/2005/proceed/papers/BASSI175.pdf>

9. Lorenz, T., Jost, A.: Towards an orientation-framework for multiparadigm modeling. In: Größler, et al. (eds.) Proceedings of the 24th international conference of the System Dynamics Society, Nijmegen (2006), <http://www.systemdynamics.org/conferences/2006/proceed/papers/LOREN178.pdf>
10. Meadows, D.H., Meadows, D.L., Randers, J., Behrens, W.: The limits to growth. Universe Books, New York (1972)
11. Meadows, D., Robinson, J.: The electronic oracle. John Wiley & Sons, Chichester (1985)
12. Morgan, M.: Experiments without Material Intervention. In: Raader, H. (ed.) The Philosophy of Scientific Experimentation, pp. 216–235. University of Pittsburgh Press, Pittsburgh (2003)
13. Morrison, M., Morgan, M.: Models as mediating instruments. In: Morgan, M., Morrison, M. (eds.) Models as mediators, pp. 10–37. Cambridge University Press, Cambridge (1999)
14. Morrison, M.: Models as autonomous agents. In: Morgan, M., Morrison, M. (eds.) Models as mediators, pp. 38–65. Cambridge University Press, Cambridge (1999)
15. Moss, S., Edmonds, B.: Towards good social science. JASS 8(4) (2005)
16. Oreskes, N.: Why believe a computer? In: Schneiderman, J. (ed.) The earth around us: Maintaining a livable planet, pp. 70–82. W.H. Freeman & Company, San Francisco (2000)
17. Parunak, H., Savit, R., Riolo, R.: Agent-Based Modeling vs. Equation-Based Modeling: A Case Study and Users' Guide. In: Proceedings of Workshop on Modeling Agent Based Systems, pp. 10–25 (1998)
18. Phelan, S.: A Note on the Correspondence Between Complexity and Systems Theory. Systemic Practice and Action Research 12(3), 237–246 (1999)
19. Sawyer, R.: Simulating emergence and downward causation in small groups. In: Moss, S., Davidsson, P. (eds.) MABS 2000. LNCS, vol. 1979, pp. 49–67. Springer, Heidelberg (2001)
20. Schieritz, N., Milling, P.: Modeling the Forest or Modeling the Trees. In: Proceedings of the 21st International Conference of the System Dynamics Society (2003)
21. Sterman, J.: Business Dynamics. Systems Thinking and Modeling for a Complex World. McGraw-Hill, Boston (2000)
22. Winsberg, E.: Sanctioning Models: The epistemology of simulation. Science in context 12(2), 275–292 (1999)
23. Winsberg, E.: The hierarchy of models in simulation. In: Magnani, L., Nersessian, N., Thagard, P. (eds.) Model-based reasoning in Scientific discovery. Springer, New York (1999)
24. Winsberg, E.: Simulated experiments: methodology for a virtual world. Philosophy of science 70, 105–125 (2003)

Algorithmic Analysis of Production Systems Used as Agent-Based Social Simulation Models

Jim Doran

School of Computer Science and Electronic Engineering, University of Essex
Colchester, CO4 3SQ UK
doraj@essex.ac.uk

Abstract. Algorithmic analysis of models is a standard tool in general, but is rarely attempted in the context of computer models for agent-based social simulation. We explore the algorithmic analysis of simulation models that take the form of **production systems** as defined in computer science. Several implemented analysis algorithms for a particular type of production system are described, including algorithms for model abstraction and for agent discovery. Examples of the use of these algorithms are given and their significance and potential considered. In particular, it is explained how an algorithm for model abstraction, developed in the context of production system models, has been successfully applied to the Iruba model of a guerrilla war, a complex multi-agent model programmed in C, a general purpose programming language.

Keywords: Model analysis, production systems, agent discovery, model abstraction, guerrilla warfare, Iruba model.

1 Introduction: Model Analysis

Agent-based social simulation relies upon the precise specification of a model in, typically, a computer programming language such as C++ or Prolog, or a multi-agent system oriented software environment such as SWARM or SDML, and then the repeated execution of the model on a computer to acquire sufficient experimental data for reliable conclusions about its behaviour to be drawn. However, as has long been recognised, this is not the only possible way to proceed (e.g. Doran and Gilbert, 1994, section 1.2.2; Teran, Edmonds, and Wallis, 2001; Gilbert and Troitzsch, 2005). A model's structure can sometimes be directly analysed to obtain information about its behaviour without its ever being executed. This can be a much faster means of obtaining properties of a large and complex model than are systematic experimental trials (see Doran, 2005). Furthermore, such analysis is arguably more natural in the sense that it corresponds more directly to the mental processes of human beings when they reflect upon dynamic systems and their behaviour.

However, it is also true that direct analysis of computer programs is very difficult, often intractable. This prompts consideration of *production systems* (in the computer science sense e.g. Wulf et al, 1981 pp 550-555; Russell and Norvig, 1995, p 314). In effect, production systems are programs in a certain class of specialised "programming languages" (CLIPS is a well known example – see, for example, Giarratano and

Riley, chapters 7-10) that have full computational power but are simple in structure and hence relatively easy to analysis. Production systems are efficient for simulation. They have often been associated with models of cognition and also with “intelligent knowledge based systems” (IKBS). We emphasise that production systems can capture complex social phenomena, including agents that learn and communicate, just as well (or as badly) as any other program.

Here we define a class of production systems, and hence production system models, and use this definition to seek to design algorithms of practical use that analyse realistic models, particularly as regards the agents that they contain, the abstractions that they support, and the interactions between agents and abstraction.

2 Production Systems

Definition: A *production system (PS)* as we define it (definitions elsewhere vary in detail) comprises three parts:

A set of variables each with its own associated possible *value set*. No variable may be associated with (i.e. bound to) more than one of its values at the same time. A set of bindings that associates a value with each variable is a *state*.

A set of rules each in the form

set of variable/value bindings => single variable/value binding

in some suitable syntax. A rule has a “left hand side” (LHS) – the binding set -- and a “right hand side” (RHS) – the single binding. Rules are not to be interpreted to specify logical entailment, but rather consequence in time. Given a state, the rules generate a successor state.

An execution procedure that matches all rule LHSs to the existing state variable/value bindings and executes in parallel all those rules that match by asserting the corresponding RHS bindings. Hence the current set of bindings is updated and a successor state generated. Bindings that are not updated by a rule execution persist unaltered to the successor state (compare *frame axioms* and their use, Russell and Norvig, 1995, page 206).

Importantly, we specify that *matching rules must be executed in parallel*. The “conflict resolution” issue prominent in alternative formulations of production systems is thus not an issue here. It is replaced by the need to avoid assertion of contradictory bindings.

The execution procedure is invoked repeatedly, and a production system thus generates a sequence of states. More abstractly, a production system implements repeated invocation of a many-one mapping over the set of possible states.

The production systems we have just defined are simple in structure and this has a number of important consequences for the programmer. *Firstly*, there are no intra-rule variables, that is, rule LHSs do not have unbound variables that are bound during

matching and the bindings then used on the RHS. The disadvantage of this simplification is that it greatly increases the number of rules typically required. *Secondly*, since matching rules are executed in parallel, the rule set of a PS must never set contradictory bindings, that is, it must never happen that as one rule sets X/a another sets X/b. *Thirdly*, it is assumed that value sets are finite and relatively small. Thus we must work with quantitative variables as if they were qualitative and design rules sets accordingly. *Finally*, pseudo-random numbers are not easily made available.

These simplifications make algorithmic analysis more tractable. The burden of conforming to them falls upon the designer of the rule set i.e. the “programmer”.

3 Two Examples of Production Systems

Definition: A *production system model* is a production system that is being used as a model.

As a first example of a PS model, consider the following small set of rules that model an aircraft landing (or crash landing!):

```
flight_status/in_air    &    engines/running    &
pilot_status/ok => flight_status/landed
```

```
flight_status/landed => engines/not_running
```

```
engines/running & fuel/4 => fuel/3
engines/running & fuel/3 => fuel/2
engines/running & fuel/2 => fuel/1
engines/running & fuel/1 => fuel/0
```

```
fuel/0 => engines/not_running
```

```
engines/not_running => altitude/ground-level
```

A plausible initial state is:

```
flight_status/in_air
engines/running
fuel/4
pilot_status/ok[contrast with
pilot_status/faulty]
```

Given this initial state, then the next state will be:

```
flight_status/landed
engines/running
fuel/3
pilot_status/ok
```

From these rules, it is easy to see that in all circumstances the aircraft must arrive at ground level with engines not running, whether or not pilot_status is initially ok.

A second example PS model, this time modelling in outline a lecture being given, is the following *standard lecture model*:

```
lecturer/speaking-well & content/knowledgeable
=> class/very-interested
```

```
lecturer/speaking-poorly & content/knowledgeable &
class/bored => class/somewhat-interested
```

```
lecturer/speaking-poorly & content/knowledgeable &
class/very-interested => class/somewhat-interested
```

```
lecturer/speaking-well & content/ignorant &
class/bored
=> class/somewhat-interested
```

```
lecturer/speaking-well & content/ignorant &
class/very-interested => class/somewhat-interested
```

```
lecturer/speaking-poorly & content/ignorant =>
class/bored
```

```
class/bored => lecturer/speaking-well
//the lecturer tries
harder!
```

```
lecturer/ speaking-poorly & class/somewhat-interested
=> class/bored
```

```
content/ignorant & class/somewhat-interested =>
class/bored
```

A possible initial state for this model is:

```
lecturer/speaking-poorly
content/ignorant
class/very-interested
```

It is reasonably easy to see by inspection of these rules that whenever *content* is *ignorant* there will be an oscillation in the class's level of interest between *somewhat-interested* and *bored*.

In this model there are, intuitively, just two agents: the lecturer and the class (collectively). In the model each of these agents comprises a single variable, which is counter-intuitively simple. In general an agent within such a model will consist of a subset of the system's variables, some of them to be interpreted as representing aspects of the agent's mental state, with associated rules.

4 Analysis Algorithms

Working from the foregoing definition of a production system, algorithms have been designed, programmed (in the C language) and tested to:

- (a) Execute a production system
- (b) Discover the prerequisite initial conditions of a specified state of a PS
- (c) Discover the stable states of a given PS
- (d) Generate “powerful” abstractions of a given PS model.
- (e) Identify the (minimal) agents in a given PS model

Algorithm (a), RUN PS, executes a PS and is relatively straightforward. In pseudo-code:

```
Repeat
{
    Match all rule LHSs against current state
    and set RHS bindings of matching rules into
    current state, overwriting existing bindings
    as required
}
```

Algorithm (b), PRECURSOR, uses “backward chaining” to find the precursor states of a particular PS model state. In the aircraft-landing model, for example, it enables algorithmic discovery of the fact that from all initial states the aircraft arrives at ground level with engines still. It also enables discovery of all initial states for the lecture model that lead to the class ending the lecture “very-interested”. The algorithm is relatively complex because of the need to handle the “frame problem” (Russell and Norvig, 1995, page 207).

Algorithm (c), STABLE STATES, finds stable states for the production system, that is states that once reached are never left.

A stable state of the standard lecture model turns out to be:

```
lecturer/speaking-well
content/knowledgeable
class/very-interested
```

Finding stable points is a standard and useful method of analysis for systems of differential equations. Indeed, systems of differential equations have much in common with production systems. This is illustrated by the fact that the standard methods of simulating systems of differential equations (e.g. Runge-Kutta) proceed by first replacing them by, in effect, a production system. In some respects production systems *generalize* systems of differential equations by permitting specification of more discontinuous relationships between variables.

5 Abstraction of a PS Model

A model of a lecture can be created at many levels of abstraction. For example, a more refined (detailed) model than the standard lecture model given above might include representations of a chairman, of slides and slide showing, and of a clock. More interesting and complex (and more obviously social) are lecture models that include representations of the mental states of the individual members of the class, of learning processes in individuals, and of interactions between class members. In fact, an *elaborated lecture model* has been created and experimented with at approaching this last level of detail (see Fig 1). It is a production system with over one hundred rules. It includes simple representations of slides being shown, of (three) class members learning, and even of two of the class members playing the little game of “paper, scissors and stone”.

```

class2 / bored game-display3 / paper => game-display2 / stone
class2 / bored game-display3 / scissors => game-display2 / paper
class2 / bored game-display3 / stone => game-display2 / scissors
class3 / bored game-display3 / stone => game-display3 / paper
class3 / bored game-display3 / paper => game-display3 / scissors
class3 / bored game-display3 / scissors => game-display3 / stone
game-display2 / stone game-display3 / scissors
                                     => game-score2 / winning
game-display2 / stone game-display3 / scissors
                                     => game-score3 / losing

```

Fig. 1. A fragment of the *elaborated lecture model*. It refers to two members (class2 and class3) of the three-member class playing “paper, scissors and stone”.

Several researchers have investigated algorithmic abstraction of models, notably Zeigler (1990), Fishwick (1995), and Ibrahim Y and Scott P (2004). Their objective has been to find automatic ways of discarding irrelevant detail from a given model, and homing in on its essential features and processes. It is standard to see models that are abstractions of a base model as the outcome of behaviour preserving homomorphisms. The difficulty is to decide what is meant by “good” abstractions and actually to find them given an original base model. We here take the view that abstraction should be applied to a production system without reference to any interpretation it may have as a model or to any particular type of structure (e.g. agents) that an interpretation may recognise within it.

More specifically, a set (or “space”) of alternative abstractions of a given base model may be generated by alternative sequences of *abstraction steps*. We specify an *abstraction step* to be the combination of two randomly chosen model variables into one, with the value set of the new combined variable obtained as the result of applying

a heuristically chosen (possibly random) many-one mapping to the cross-product set of the value sets of the two original variables. The cardinality of the value set of the new combined variable is normally much less than the cardinality of the cross-product set of the two original variables.

In the work reported here (algorithm (d) above, ABSTRACTION) these abstraction steps are used within a heuristic locally guided search (also called *hill-climbing search*, see Russell and Norvig, 1995, 112). Successive abstraction steps are chosen and appended to the existing sequence of adopted abstraction steps. The choice between alternative abstraction steps is made by always selecting the abstraction step that is heuristically assessed as leading to the least indeterminacy.

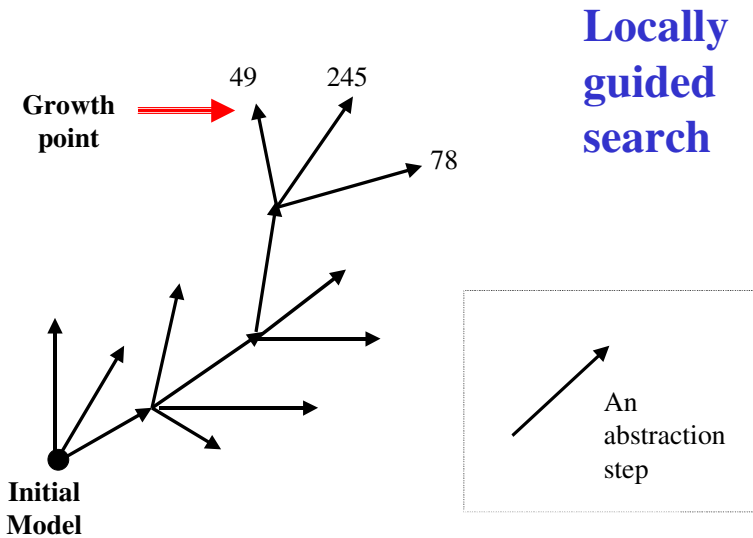


Fig. 2. Locally guided search through the space of model abstractions

The *indeterminacy* of an abstraction step is essentially measured as the number of indeterminate (ie one-many) abstracted state successions to which the proposed abstraction step (and preceding step sequence) gives rise from a sample of basic model state trajectories generated for the purpose. The rationale for this approach is that we accept a “class”, say, as an abstraction of a set of “students” if we find the concept of a class useful as a means to simple and sound predictions and explanations.

In outline pseudo-code, the algorithm ABSTRACTION is:

```

Input the base model

Until no further abstraction is possible
{
    Generate a set, S, of alternative abstraction
steps

```

```

    Select and adopt the least indeterminate
    abstraction step from S by testing each pos-
    sible abstraction step on an ad hoc sample of
    state trajectories.
  }

```

Applying ABSTRACTION to the standard lecture model yields the following simple abstract model that highlights the possible oscillation in the class’s level of interest between *somewhat-interested* and *bored*.

```

V/x  => V/x
V/p  => V/q
V/q  => V/p

```

This highly abstracted model has just one variable, and it is apparent that either its state is constant (V/x) or it alternates between V/p and V/q.

Finding powerful abstractions potentially provides insight into the essential and significant behaviour of a model and of the phenomena that it models. The existing algorithm can generate a range of abstract models from the standard lecture model listed above and highlights key behaviours. Experimental application of ABSTRACTION to the *elaborated lecture model* is ongoing.

To reiterate, the major difference between this approach to model abstraction and those mentioned earlier is that we define abstraction operations at the “fine-grained” level – at a level below that at which such macro-phenomena as agents are naturally considered -- and that we have deployed a standard heuristic search procedure to locate good abstracted models in the space of possible abstracted models.

Indeterminacy

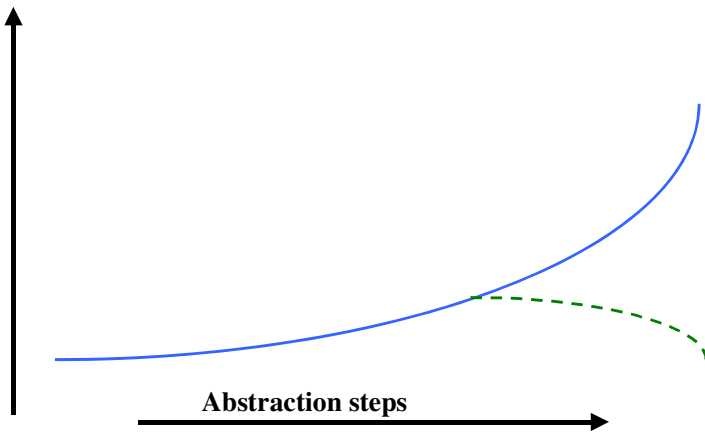


Fig 3. Poor abstractions increase indeterminacy (the blue line). Effective abstractions reduce indeterminacy (the dashed green line).

6 The Iruba Model: A Full Scale Example

Crucially, it turns out that the method of model abstraction just described is applicable not merely to models in PS form, but to a much wider range of programmed models. To explain and illustrate how this is achieved we describe the Iruba model of a guerrilla war and the way in which this method of abstraction has been applied to it.

The Iruba project at the University of Essex (Doran, 2005, 2006) is following standard agent-based social modelling procedure to construct and experiment with a holistic generic model of a guerrilla war sufficiently realistic to offer new insights into guerrilla war dynamics or to further develop existing insights. In the model agents correspond to guerrilla bands, regime bases or outposts, and to headquarters on each side. A particular objective is to establish sets of conditions expressed in terms of the model's parameters and structures that guarantee that an insurgency will succeed or, alternatively, will fail.

The Iruba model has been made broadly realistic having regard to the relevant literature. In particular, the model is loosely based on (extensive descriptions of) guerrilla wars that took place in the last century in Ireland (1919-1921) and in Cuba (1956-1959), with some further features drawn from the Arab Revolt against the Turks in the Hejaz (1917-1918) towards the end of the First World War.¹ A reliable source for these conflicts is Beckett (2001) but there are many others. The most important structural and behavioural concepts used in building the Iruba model are drawn from the Irish insurgency: near autonomous regions (on an island) with only limited central control; mobility notably in the form of "flying columns"; limited weaponry; the importance of terrain; and the importance of ideology and popular support.

The Iruba model is structured as a network of 32 relatively autonomous regions that vary in terrain and population. The population of a region provides a (finite) recruitment pool for both insurgents and regime forces. Initially the regime forces are relatively numerous, distributed in bases over the island, and relatively static, whilst the insurgents are small in number, localised, mobile and hard to find. As indicated, computational agents represent guerrilla cells/bands and regime bases and insurgent and regime headquarters. Attacks take place within regions following simple rational strategies. For example, a guerrilla band may attack a poorly defended regime base, with the outcome dependent upon terrain, relative numbers and weaponry, and random factors. A successful attack may well lead to capture of weapons. Movement of insurgent or regime forces between neighbouring regions takes place under appropriate conditions. For example, neither the forces that are moved nor those that remain behind are left at significant risk (but see later – hyper-mobility). Recruitment to insurgents and to regime forces (and defection from regime forces) reflects the numbers and attitudes of the so far uncommitted general population of the region in question. This population will partially support the insurgents, and will partially be aware of the insurgency, depending upon the conflict history in that region. These two "attitude" variables and their use are intended to go some way towards capturing the dynamics of population opinion and its impact upon the course of the insurgency.

¹ In each of these examples, the insurgents proved (more or less) successful. However, as stated, the structure of the Iruba model equally allows for regime success.

The core cycle of the model may be expressed in outline pseudo-code as:

```
Repeat
  Attacks and their impact
  HQ decisions
  Recruitment
  Force movement
Until termination
```

As indicated above, a degree of central control by “headquarters” agents is possible for both sides. If an insurgency grows, regime forces may be concentrated into regions where the insurgency is at its strongest. Furthermore, faced with a dangerous insurgency the regime may take “all out” measures (comparable with, for example, the so called “Salvador option”).² On the other side, in appropriate circumstances the insurgents may be switched into “hyper-mobile” mode (comparable with the use of “flying columns” by the IRA in Ireland) and/or an “all out” attack across a range of regions or even the entire island, may be triggered (compare the Tet Offensive in the Vietnam war).

Victory in this model is a matter either of insurgent annihilation, or of the insurgents achieving numerical superiority and hence, by assumption, political power. At several points the model invokes chance factors (using a pseudo-random number generator) so that the success or failure of an insurgency may vary with the pseudo-random number stream seed even if all other model setting are the same.

The Iruba model has been implemented in the C programming language. In total the model contains well over a hundred variables, of diverse types. Although some model variables (e.g. population support for insurgents) are updated by simple mathematical relationships, many aspects of the model structure are more complex. For example, agents (guerrilla bands, regime bases and HQs) are essentially expressed as sets of conditional rules.

7 Initial Abstraction Results from the Iruba Model

Within the ABSTRACTION algorithm the Iruba model in its entirety is, in effect, repeatedly executed as a subroutine in order to generate multiple model trajectories to which alternative sequences of mappings may be applied. A search is conducted through the space of abstraction, or compound mappings, of the Iruba model (currently a subset only of the model’s variables) as indicated earlier in Figure 2.

It has proved quite possible to obtain meaningful abstractions of the Iruba model in under a minute on a standard PC. For example, initial results highlight that total insurgency size tends to zero (insurgent defeat) or to a maximum (insurgent victory), excluding oscillation. This result is obtained although there is no variable for total insurgency size in the Iruba model itself, which represents the insurgency as distributed over multiple regions. More difficult is to tune the heuristic search process to select automatically the most powerful abstracted models, and this is a focus of current experimentation.

² For a discussion of the “Salvador option” see Michael Hirsh and John Barry, NEWSWEEK, Jan 10th, 2005.

8 Agents in a PS Model

Can a PS model contain (representations of) agents? The answer is surely yes (see earlier). With each agent in the model interpretation there may be associated a subset of the model's set of variables, some of them reflecting presumed "mental" and "emotional" attributes. A subset of the rules may also be associated with a particular agent. Other variables and rules will represent interactions and the agents' collective environment. It is thus natural to ask the following question. Can the agents that a model creator(s)'s interpretation of the model suggests to be present within it, be detected by an "objective" algorithm that has no access to the interpretation?

For example, in the standard lecture model (above) there are intuitively just two agents, the lecturer and the class. But would an objective analysis of the production system itself reveal these two agents. In other words, do the agents in an agent-based model have a real computational existence beyond the interpretation read into the model by its designer and users? Should they? Can there be other agents in the model that its creator is not aware of, that is, which are in some sense emergent from the model's structure? We suggest that the ability to discover the minimal agents formally present in a PS model is useful. It enables comparison with the set of agents the model's designer *conceives* as present, possibly leading either to modification of the model or of the model design concept.

However, defining in computational terms just what is an agent in the context of a PS is difficult, and widely recognised to be so. Standard textbook definitions are either too specialised or too ambiguous to offer much help³. We define a *minimal agent* in a production system by reference to the notion of "influence" as follows (adapted from Doran, 2002):

One variable in a production system is said to *influence* another when there exists at least one rule that refers to the former on its LHS and binds the latter on its RHS.

A *minimal agent* is a non-empty subset of the variables of a production system. These variables must each be one of *input*, *output*, or *internal* variables where the meaning of these agent variable types is defined by reference to influence.

- An input variable is influenced only by variables external to the agent and influences only output variables and/or variables internal to the agent.
- An output variable is influenced only by input variables and/or internal variables of the agent and influences only variables external to the agent.
- An internal variable is influenced by, or influences, only other variables of the agent i.e. internal, input or output variables.

³ E.g. "An agent is a computer system that is situated in some environment and that is capable of autonomous action in this environment in order to meet its design objectives" (Wooldridge, 2002, p.15).

An agent must contain at least one input variable and one output variable, and must include ALL variables that are internal consequent on these input and output variable sets.

The rationale for this definition is to express a minimum intuitive requirement for an agent: an agent “perceives”, “decides”, and “acts”. A minimal agent according to this definition may be very simple, no more than an input and an output, or it may be highly complex and “cognitive” – neither is precluded. Typically, the agents (or representations of agents depending on your viewpoint) in current agent-based social simulation studies rarely comprise more than a few rules and variables. Notice that according to this definition two agents may or may not be disjoint – consider that the audience in a lecture is both a single agent and made up of many agents. Agents may also be nested.

Algorithm (e), FIND AGENTS, is able to discover the agents, in this sense, in a production system. However, at present it is limited to finding agents that have exactly one input variable and exactly one output variable. In outline pseudo-code FIND AGENTS is as follows:

```

Foreach possible pair of variables V_input, V_output

{Find the set of variables S1 directly or indirectly
influenced by V_input
Find the set of variables S2 that directly or indirectly
influence V_output
If V_input is not itself a member of S1
and V_output is not itself a member of S2
then, provided the intersection of S1 and S2 is not null,
the union of S1 and S2 is an agent
}

```

When applied to the standard lecture PS model given earlier, this algorithm finds no agents as is to be expected. There are only three variables in the model, and its structure is too simple for it to have agent content. However, when the model is developed so that explicit perceptual input and appearance output variables are associated with the class (for details see Appendix A), then the three variables now constituting the class are indeed recognised as an agent, as also are the variables that together constitute the “environment” of the class.

Using the above definition it is possible to make precise, for example, how two or more minimal agents may be viewed as a single agent (compare Bosse and Treur, 2006). It is an open question how best to strengthen the definition to render the agents it delimits more clearly cognitive whilst keeping the definition sufficiently precise for purpose.

It is natural to ask what agents FIND AGENTS will discover when this algorithm is applied to the Iruba model. The question is all the more interesting because whilst it is natural to see insurgent and regime *groups* as agents (as suggested earlier), from a more spatial perspective it is also natural to view *regions* as agents. There is a substantial difficulty in practice. The Iruba model is not in PS form. Therefore, before

FIND AGENTS can be applied the Iruba model must be converted to PS form. More precisely, the influence “skeleton” of the Iruba model must be extracted from the code (by hand or otherwise) and made available to the FIND AGENTS algorithm. This investigation is ongoing.

9 Agents and Abstraction

It is natural to ask under what conditions agents existing in a PS model are preserved under the abstraction process. A plausible conjecture is that if abstraction steps are constrained always to use pairs of variables that are either both within or both outside every agent in the model, those agents will either be preserved or reduced to a single variable (thereby ceasing to be an agent) by abstraction. Agents will never be introduced by the abstraction step. However, were two variables one within and one outside an agent to be combined, then this would be likely to destroy agenthood. Hence it seems that there do exist conditions under which abstraction preserves agents until they are reduced to single variables. The significance of this insight is that if the original set of agents in a model is an important part of our view of the target system to which the model relates, in practice we may wish an abstraction process to have this agent-preserving property. More generally, we can ask for each type of macrostructure that we find of interest in the PS model under what conditions it is preserved under abstraction.

10 Discussion and Conclusions

This paper has addressed automatic analysis of a model to gain insights into its behaviour. Few would dispute that in principle such analysis can be useful. But the practical value of the lines of attack presented depends not just upon the computational feasibility of the algorithms discussed, which has been shown at least for simple cases, but also on the possibility of creating non-trivial PS models in the first place. There is no doubt that constructing PS models of the type we have considered is potentially laborious and intricate, primarily because there are no intra-rule variables, and because the value sets must be small. Some systematisation of the model creation process looks essential. Models with “spatial” structure may be relatively easy to construct. There are other options. It may be possible to specify models in a higher-level language and then automatically compile them down to the simple PS level considered here. If this proves effective, then the burden upon the PS model creator(s) will be correspondingly reduced. We conclude that to the extent that valid PS models can be created, and analysis algorithms of the type described above applied to them – and analysis of large models comprising hundreds of rules is quite tractable – then these techniques promise to be useful alongside (not, of course, instead of) the standard methods of agent-based social simulation.

However, there is another possible road to practicality – to develop algorithms in the context of PS models, but then to extend their model domain. Thus, as reported earlier, it has been shown that the ABSTRACTION algorithm, in particular, is usefully applicable to a full-scale model (the Iruba model) written in a general purpose

programming language. We may therefore see production system models as a testing ground for insights and algorithms that can then sometimes be recast in fully general form.

Although this paper has primarily been about methods, there are some more general epistemological points to be made. The first point is to reiterate that more can be done with a simulation model that just to run it on a computer to generate simulation traces. A simulation model can be analysed, including algorithmically. Analogues of mathematical styles of analysis are possible. It is implicit in these assertions that a model is really no more than a knowledgeable description that admits of certain types of reliable inference procedure. This paper has concerned ways in which to extend the set of available inference procedure for a certain class of model.

Secondly, discovery of abstract models is akin to concept discovery. Concepts are abstractions from sets of instances. It follows that model abstraction process may also be viewed as concept discovery processes.

Thirdly, the inference processes we apply to models need not (and often should not) distinguish between those parts of models that correspond to agents (whether or not the agents are in some sense “social”) and those that correspond to non-agents.

Finally, the difference between an “agent-based model” and a “diagnostic expert system” (ES) or “intelligent knowledge based system” (IKBS), in the traditional sense, is much less substantial than it at first appears. Consider an agent-based model used to determine whether or not a guerrilla war will be won by insurgent or regime forces (looking back to the Iruba project). The simulation model must incorporate domain knowledge and use it to determine the likely outcome of the war just as a medical expert system might use domain knowledge to predict the outcome of a course of treatment for some disease for a particular patient. And, indeed, there are well-known medical expert systems (e.g. CASNET/GLAUCOMA, Weiss et al., 1978) that use simulation as a component of their processing. To relate agent-based modelling to expert systems in this way is unusual but offers insights that may prove helpful.

References

1. Beckett Ian, F.W.: *Modern Insurgencies and Counter-Insurgencies: guerrillas and their opponents since 1750*. Routledge, London (2001)
2. Bosse, T., Treur, J.: Formal Interpretation of a Multi-Agent Society As a Single Agent. *JASSS* 9(2) (2006), <http://jasss.soc.surrey.ac.uk/9/2/6.html>
3. Doran, J.E.: Agents and MAS in STaMs. In: d’Inverno, M., Luck, M., Fisher, M., Preist, C. (eds.) *UKMAS Workshops 1996-2000*. LNCS (LNAI), vol. 2403, pp. 131–151. Springer, Heidelberg (2002)
4. Doran, J.E.: Iruba: an Agent Based model of the Guerrilla War Process. In: Klaus, G., Troitzsch, F. (eds.) *Representing Social Reality, Pre-Proceedings of the Third Conference of the European Social Simulation Association (ESSA)*, Koblenz, September 5–9, 2005, pp. 198–205 (2005)
5. Doran, J.E.: Modelling a Typical Guerrilla War. In: *Proceedings IEEE DIS 2006 Workshop (Knowledge Systems for Coalition Operations strand)*, Prague (June 2006)
6. Doran, J.E., Gilbert, N.: *Simulating Societies: an Introduction*. In: Gilbert, N., Doran, J.E. (eds.) *Simulating Societies*, pp. 1–18. UCL Press (1994)

7. Fishwick, P.A.: Simulation Model Design and Execution. Prentice-Hall, Englewood Cliffs (1995)
8. Giarratano, J., Riley, G.: Expert Systems: Principles and Programming. PWS-Kent Pub. Co., Boston (1989)
9. Gilbert, N., Troitzsch, K.: Simulation for the Social Scientist, 2nd edn. UCL Press, London (2005)
10. Hart, P.: The I.R.A at War, Oxford, pp. 1916–1923 (2003)
11. Ibrahim, Y., Scott, P.: Automated Abstraction for Rule-Based Multi-Agent Systems. In: Proc. ESSA 2004, Valladolid, Spain (2004)
12. Russell, S., Norvig, P.: Artificial Intelligence, a modern approach. Prentice-Hall, Englewood Cliffs (1995)
13. Teran, O., Edmonds, B., Wallis, S.: Mapping the envelope of social simulation trajectories. In: Moss, S., Davidsson, P. (eds.) MABS 2000. LNCS (LNAI), vol. 1979, pp. 229–243. Springer, Heidelberg (2001)
14. Weiss, S.M., Kulikowski, C.A., Amarel, S., Saffir, A.: A model-based method for computer-aided medical decision-making. Artificial Intelligence 11, 145–172 (1978)
15. Wooldridge, M.: MultiAgent Systems. Wiley and Sons, Chichester (2002)
16. Wulf, W.A., Shaw, M., Hilfinger, P.N., Flon, L.: Fundamental Structures of Computer Science. Addison-Wesley, Reading (1981)
17. Zeigler, B.P.: Object-Oriented Simulation with Hierarchical, Modular Models: intelligent agents and endomorphic systems. Academic Press, London (1990)

Appendix

There follows the rule set for the standard lecture model with added structure. This structure takes the form of a variable *classp* that mediates perceptions made by the class and a variable *classa* that mediates the external appearance of the class. The three variables *classp*, *classa* and *classa* are then collectively recognised as an agent by the algorithm FIND AGENT. The variables *classa*, *lecturer*, *content* and *classp* are also collectively recognised as an agent (in effect, the agent’s “environment”), itself a sequential combination of two agents.

```
lecturer / speaking-well & content / knowledgeable =>
classp / well-k
```

```
classp / well-k => class / very-interested
```

```
lecturer / speaking-poorly & content / knowledgeable =>
classp / poorly-k
```

```
classp / poorly-k & class / bored =>
class / somewhat-interested
```

```
classp / poorly-k & class / very-interested =>
class / somewhat-interested
```

```
lecturer / speaking-well & content / ignorant =>
classp / well-i
```

classp / well-i & class / bored => class / somewhat-
interested

classp / well-i & class / very-interested =>
class / somewhat-interested

lecturer / speaking-poorly & content / ignorant =>
classp / poorly-i

classp / poorly-i => class / bored

class / very-interested => classa / very-interested

class / somewhat-interested => classa / somewhat-
interested

class / bored => classa / bored

classa / bored => lecturer / speaking-well

classp / poorly-k & class / somewhat-interested =>
class / bored

classp / poorly-i & class / somewhat-interested =>
class / bored

classp / well-i & class / somewhat-interested => class
/ bored

classp / poorly-i & class / somewhat-interested =>
class / bored

The Nature of Noise

Bruce Edmonds

Centre for Policy Modelling, Manchester Metropolitan University
<http://bruce.edmonds.name>

Abstract. The idea of noise is now widespread in many fields of study. However to a large extent the use of this term is unexamined. It has become part of the practice of science without entering to a significant extent as part of its explicit theory. Here I try to produce a clearer and more coherent account of the term. I start with a picture of noise from electrical engineering. I then generalise this to the widest conception: that of noise as what is unwanted. A closely related conception is noise as what is unexplained. A particular case of this later usage is where a source of randomness can be used to stand-in for this residual. I argue that noise and randomness are not the same. I explore the possible relation between noise and context, and propose a new conception of noise: namely that noise is what can result from an extra-contextual signal. I finish with an application of the analysis of noise to the relation of determinism and randomness.

Keywords: Noise, relevance, modelling, explanation, randomness, residual, context, determinism, philosophy, science.

1 Introduction

The idea of noise now plays a prominent (if subsidiary) part in many fields of study; it is casually mentioned in numerous papers as if its nature was well established. However to a large extent the term is not examined, but simply used. In other words, it has become part of the *practice* of science without entering to a significant¹ extent as part of its explicit *theory*².

The Oxford English Dictionary presents the following account (after a list of descriptions of more mundane uses referring to the older, non-technical uses of the word noise):

“11. a. In scientific and technical use: random or irregular fluctuations or disturbances which are not part of a signal (whether the result is audible or not), or which interfere with or obscure a signal; oscillations with a randomly fluctuating amplitude over a usually continuous range of frequencies. Also in extended use: distortions or additions which interfere with the transfer of information....

¹ The exception being in Electrical Engineering where some attention has been paid to this area, however this is specific to where it is possible to include some properties of the noise within an explicit (statistical) model. For more on this see the second section.

² (Habermas 1963), this seems to relate to the distinction in (Cartwright 1983) on bridging rules and theory.

b. In non-technical contexts: irrelevant or superfluous information or activity, esp. that which distracts from what is important.”

Thus we have a veritable “hairball” of related ideas, including: randomness; irregularity; disturbance; interference; obscurity; non part of a signal; distortion; addition; irrelevance; superfluity; and distraction applying to either a signal or to “what is important”. I think this accurately represents the term as it is used in technical papers (both in terms of content and in its vagueness). In this paper I try to produce a clearer and more coherent account of the term – an account that moves towards a more general account of noise and, in particular, an account that will be of use to computational modellers.

I start with a picture of noise from electrical engineering, since this seems to be where the concept arose first in academic papers³ and so frames much of the scientific thinking on the subject. I then generalise this picture to the widest conception: noise as that which is unwanted, which relates to our aural experience. A closely related conception that has developed in the scientific literature is noise as what is unexplained – the residual after what can be explained (or modelled) is factored out. A particular case of this later usage is where a source of randomness can be used to stand-in for this residual. This strategy has led some to almost identify noise and randomness. However I argue that noise and randomness are *not* the same and thoughtlessly conflating them can result in misleading assumptions. I explore the possible relation between noise and context, and propose a new conception of noise, namely that noise is what can result from an extra-contextual signal. I claim that this is not only a psychologically plausible account of its origin (and hence relates well to common usage) but is also a useful way of conceptualising it. I finish with an application of the analysis of noise to the relation of determinism and randomness.

2 A Picture of Noise in Electrical Engineering

The close correspondence between electrical and acoustic phenomena allows the transfer of the term “noise” from an audible description to that which describes elements in the electrical phenomena that generates the sounds via a loud speaker. When you hear reproduced sound that is generated by electrical apparatus and compare it to the original sound, it has changed: a “hiss” has been added that sounds like (and acts like) noise. Thus noise becomes a technical term by analogy, however the way it was done is important and seems to have influenced the idea’s subsequent development.

The focus of much work in electrical engineering is the manipulation of identifiable electrical *signals*. These signals are the patterns or values encoded by the properties of a real electric current or field. To be precise, the signals are the intended patterns or values that are represented in actual (i.e. implemented or observed) flows of electrons and the forces between electrons. This distinction between *electrical signal* and *electrical phenomena* is important because it marks the shift from a scientific point of

³ The Oxford English Dictionary’s first quote from the academic literature is from the Telegraph and Telephone Journal in 1923 (vol. 9 pp. 119/2) “The variations in noise were plotted, and their effect at times was to reduce the intelligibility to 20 or 30 per cent.”. Obviously, the non-technical senses of the word are much older.

view, where one is trying to discover the properties of the observed world, to an engineering perspective, where one is trying to manipulate parts of the world to obtain a desired effect. Of course, there are no well-defined boundaries between these two approaches, and in practice many who work with electrical phenomena will, at different times and for different purposes, swap between scientific and engineering viewpoints. It is noticeable that ‘noise’ enters the scientific literature exactly when we were able to control these effects sufficiently in order to be able to manipulate electrical phenomena in intended ways.

In electrical engineering you have a set of *inputs* and *outputs* that pass through a set of (usually well-defined) circuitry. The aim is to effect an intended transformation upon the inputs, which become the outputs. The transformations are implemented by (on-the-whole carefully manufactured) components such as transistors, diodes etc. The heat and nature of these devices means that the intended transformation is imperfect. For example an amplifier may be supposed to evenly increase the amplitude of a signal but otherwise leave it unchanged, but in reality will also distort it, add ‘hiss’ that is audible when broadcast etc. This situation is illustrated below in Fig. 1.

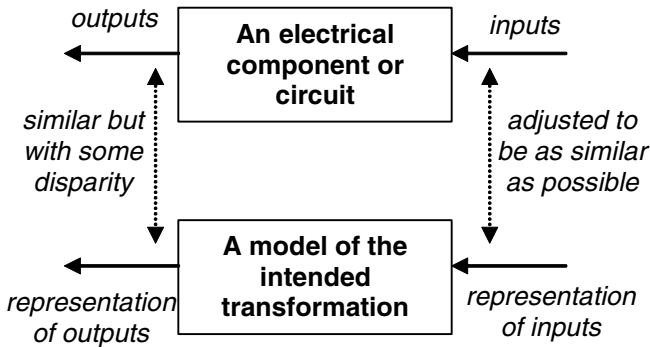


Fig. 1. The disparity between the actual and ideal outputs of a circuit

In electrical engineering there are relatively good models for the action of such components (or conversely, we have learnt to manufacture components so that they are well represented by our models). Often these models not only specify the principal transformation effected by a component but also give some information about the nature and extent of the imperfections, particularly the result of heat (and other entropy producing processes). This kind of understanding goes back to (Einstein 1905) which seems to pre-date the technical use of the word “noise”. Although the detail of this noise is unpredictable, it is particularly well characterised in terms of its general properties and is closely mimicked by a random statistical process – we call this ‘white’ noise. Thus many models of electrical components *reify* this as a distinct ‘noise term’, which may be otherwise undefined, or may have a defined probability distribution and related its magnitude to conditions such as temperature. The source of this particular kind of imperfection is relatively well understood, predictable and perceptible – it thus is natural to conceptually separate it out and label it (and the

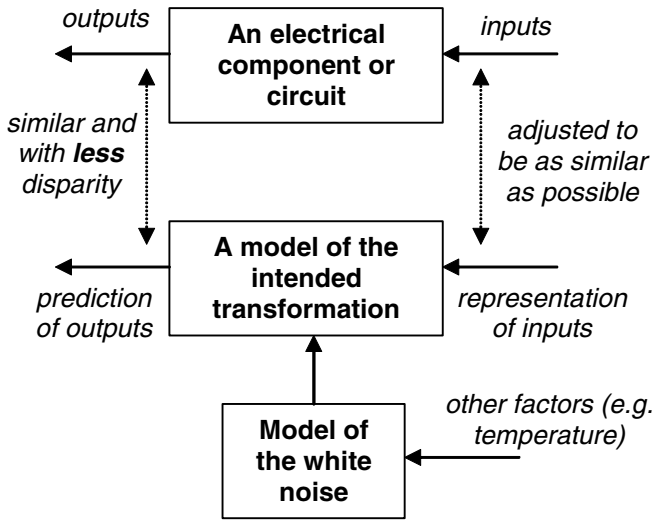


Fig. 2. The smaller disparity when noise is reified and explicitly represented

analogy with sound suggests the label ‘noise’). Thus the notions of the intended signal and the actual phenomena diverge. I illustrate the result of this process of abstraction in Fig. 2.

This shows that it can be undoubtedly useful to think of noise as a separate and identifiable extraneous factor that is ‘added’ to the signal to obtain a more realistic prediction of the results. The combined model is composed of two submodels: the original and the model of the residual⁴. This coincides with the ‘shift’ from a *scientific* viewpoint of such circuits to an *engineering* viewpoint. As Wong (2003) puts it:

“Particularly, to the physicist the noise in an electronic system represents a practical manifesting of the phenomena described by statistical mechanics, and an understanding of its practical consequences helps to illuminate and clarify some concepts of the physical theory; to the electronics engineer, noise is a constraint of the real systems, but a better understanding of its physical origins helps the engineers to minimize its effects by informed and careful design.”

Within electrical engineering there is a continuing awareness that the randomness is just a model for the noise and that this model has its limitations. Pierce makes it clear that the noise is the primary descriptor and randomness part of the model when he says (1956):

“Many sorts of electrical signals are called noise ... many engineers have come to regard any interfering signal of a more or less unpredictable nature as noise. ... The theory of noise presented here is not valid for all signals or phenomena which the engineer may identify as noise.”

⁴ Alternately one can consider the conceptual and phenomenal models as separate (Hughes 1997).

This sort of practice is now quite common in many fields, often indicated by the appearance of a ‘noise’ term⁵ in equations. However it should be remembered that the above case is quite a special one. This particular kind of distortion found in electrical engineering is particularly well-understood, predictable and separable. It is also only one particular kind of distortion that can occur, albeit often a dominant one – other kinds such as resonance, interference, distortion etc. also occur and are more difficult to predict/model partly because they are less separable from an intended (or significant⁶) *pattern*. Mimicking this step of using a single reified source called noise to stand for a general disparity between explained and actual outcomes is likely to be much less useful in many other domains.

Some of what makes the electrical engineering case special (i.e. justifies the attribution of ‘noise’ to error) can be summarised as follows:

- The input signal and the intended results are known;
- There are good predictive model of the intended transformations involved;
- The whole system is engineered in a controlled way – the composition of the parts is explicitly known;
- The principal disparity between ideal and actual is well-understood;
- Many of the disparity’s characteristics are predicatively modelled as a separate part of the model;
- This disparity can be characterised numerically so that the total disparity for a system can be estimated mathematically;
- This disparity is easy to identify and even directly perceive.

These conditions have made the reification of noise as kinds of randomness a useful, if minor, field of study⁷. Analogous conditions are evidently not true for many other domains of study, and yet this idea of a separable source of noise being mixed in with a significant signal has become a common one. In the next section I consider a generalisation of the concept of noise to include the wider cases.

3 Noise as Unwanted Interference

Noise can be seen as a largely a negative idea – as humans we are often interested in specific sound *signals* in terms of communication, music etc. and other sounds may make the task of identifying these signals more difficult. The difference between a set of measurements and an ideal is often attributed to ‘noise’ – in this case it is precisely what is *not* the ideal. This is the older use of the term, that precedes its entrance into the academic literature. If you are trying to listen to the radio and there are busses travelling outside or even people talking you may lump these together as ‘noise’, since

⁵ Although frequently this has no predictive function but is merely an apology for the disparity between the prediction and what is observed, i.e. it does not predict the extent or nature of the disparity (as is frequently the case in electrical engineering) but is there as an admission of incompleteness. In these cases it is more correct to call this an ‘error term’.

⁶ Of course, what is considered significant is a movable feast – but this is the point: attributions of ‘noise’ are results of relevance decisions, even if these are implicit.

⁷ Useful collections of sources about electrical noise are: (Gupta 1977, Ritter 2003).

they are extraneous to what you were concentrating upon. One pair of people talking may be noise from the point of view of another pair of people talking and *vice versa*.

Noise here, actually or potentially, gets in the way of a perception or observation. A signal that has no chance at all of interfering with another is simply an irrelevant signal. Even if my neighbour's car generates a lot of sound inside its engine as a side effect of its operation, it is not noise if its silencer is such that this never escapes so as to disturb me. Thus a silencer on a car *eliminates* noise, even if the sound levels *inside* that engine remain the same. It is critical that if one counterfactually *imagines* the sound as escaping then it is natural to think of this as potential noise.

This conception of noise has, by analogy, now extended beyond aural noise – it is now commonly used as a generalisation of this idea. Thus an electric drill might be said to generate electrical ‘noise’ which might interfere with the reception of my TV. The essential aspects of this generalised conception seem to be that:

- there is an identifiable source of the ‘noise’
- that interferes with the reception of a target signal
- such that the noise is not intrinsic (or essential) to the signal and its transmission.

A non-electrical example of this kind of noise is interferences in gene expression (Blake et al. 2003).

4 Noise as Unmodelled Residual

To a modeller, noise is that which prevents one from modelling a process with complete accuracy. In a way the noise is something that is interfering with the modelling results, and so can be seen as analogous to the characterisation of noise in the previous section. Thus an econometric model may be composed of an identified trend plus a noise term to ‘account’ for the deviations from this trend.

This is a presumptuous use of the term – it implies that the ‘noise’ is something which may mask the hypothesised trend but is judged to be ultimately irrelevant to it. It thus suggests that the disparity is not simply due to modelling error, i.e. that there is not an accessible better model that would match the known evidence more closely.

Nonetheless attributing modelling error to noise is sometimes appropriate – if you count the number of children at a children's party where children are not allowed to enter or leave and you get a different count each time then one can safely attribute the error to measurement “noise” (expressed in the chaotic movement of the children) rather than fundamental model error (the model being that there is a fixed integral number, n , of children in the room). You can sensibly model the sequence of counts as a fixed constant plus a noise term (which might usefully be given a random nature).

Whatever the purpose of the modelling, a discrepancy between the model outputs and target data gained from measurement of the modelling target is an indication that there is something in the target+measurement processes that is not captured by the model. In other words, the data itself is another model, a “data model” (Suppes 1962). If one has some good reasons to attribute this discrepancy to some factors that are not relevant to (in the sense of “arbitrary with respect to”) the model and its purpose, then it can be properly called noise.

When modelling complex phenomena (e.g. social phenomena) it is inevitable that one is not going to be able to capture all aspects of one's target in any single model and hence one would expect that there will be a disparity between a model's outputs and the corresponding data obtained from the target. The key question is whether this discrepancy is due factors that are relevant to the modelling task or originate from a process that is independent and (for the current purposes) irrelevant.

Conflating noise and fundamental model error can be very unhelpful. I will illustrate this with an example.

Example 1: The estimation of insurance claims

This case concerns the estimation of the frequency of events in the insurance industry. Before 1990, the distribution of insurance claims was thought to be roughly normal, that is when one plotted a the frequencies of periods with different levels of claims the result fitted a normal curve well except for a few periods with a very high level of claims. However these few cases were discounted as due to particular circumstances and hence unmodellable. This model of a normal distribution plus some arbitrary events was used to set premiums. However more recently a series of models which explicitly includes and predicts such "extreme" events have been re-discovered and applied (Black 1986, Andersen and Sornette 2001).

Many social models attempt to use statistical models to separate out social trends from the 'noise', under the assumption that human actions can be represented *en masse* as trends plus arbitrary actions that tend to cancel each other out (in large samples). If that 'noise' is supposed to be effectively random, then this is a hypothesis that can be tested, since in increasingly large samples the noise will not grow as fast as the signal and will thus fall as a proportion of the population size (something that is not obviously true of many social systems, e.g. stock markets). This is called the "law of large numbers". This may be justified as in (Dichev 2001) but in other cases it is not, as demonstrated by the models of (Kaneko 1994, Edmonds 1999a).

Indeed we are so used to the idea of separating out a signal from added random noise that we forget what a special case this is, as (Van Kempen 2001) put it:

"The amazing thing therefore is not that they [fluctuations] give rise to irregular phenomena. The amazing thing is that a collective behavior emerges, which is regular and can be described by general law..."

5 Noise and Randomness

In the special case where we are trying to model a particular data-generating process, we may well get to the stage where we have included in the model any systematic pattern we can detect in the data. In such a case what is left is precisely what we can not systematically model, namely a residue that is effectively unpredictable⁸. This

⁸ Or, to put it in a weaker form: effectively unexplainable.

residue is, by construction, not modelled. However it might be possible to adjust the model so that the secondary characteristics of this residue are modellable. For example, although each point is unpredictable, the statistical moments of an increasingly large sample of this data may converge to particular values. These values can be used to construct a sort of model of these residuals – a statistical model. Such a statistical model is composed of a particular probability distribution, from which values can be randomly (or pseudo-randomly) generated. This random series of values is not the same as the unmodelled residual but 'looks' the same to us as modellers – each has the same known distribution and the same unpredictable content within this. If we deem that the unpredictable residue is not relevant then we can say that in all relevant aspects the random series and the residue are the same. Thus it is that we often think of this unmodellable residue as random – they *look* the same to us and (by construction) from the point of view of the model their detail is equally irrelevant⁹.

Such a perspective might lead one to conflate such noise and randomness and identify noise *as* randomness. However, this is simply to confuse a model (randomness) with what is modelled (noise). That they are not necessarily the same as the following examples show.

Example 2: The interspersed coded-messages

Imagine two different messages (strings *a* and *b*) that are both encrypted (to strings *x* and *y*) by different people (*A* and *B* respectively) so that each appears to be a random sequence to anyone but the person who encoded it. Then the two sequences are interspersed to form a single sequence (*c*). To *A* the part of *c* that is *y* is just noise, there is nothing modellable about *y*, only *x* has meaning. To *B* the part of *c* that is *x* is the noise whilst the *y* part has meaning.

The part that is noise for one person is the signal for the other and *vice versa*. Thus either randomness is a relative concept or it is different from noise (which in this case is clearly relative).

One might think that such a type of sequence would not result from any natural process, but this is only an assumption. If it is possible to make an unguessable binary sequence, then it is possible to implement such a data-generating process as a program on a computer. Thus if we had such a computer attached to a Geiger counter and a small lump of radioactive material (to generate the random input), and we always plug more memory into it as required, we could produce such a sequence. Thus although highly contrived such an example is not beyond the bounds of possibility for a natural process.

Example 3: The sabotaged random number generator

Imagine a company that sells good quality random sequences, that is sequences with well-defined long-term statistical properties but with no

⁹ There are a few approaches that move towards defining randomness as the unmodellable residual, e.g. (Compagner 1991). It also turns out that patterns that are not compressible by a Turing Machine (Kolmogorov 1965) passes many of the traditional tests for randomness.

guessable patterns in the detail at all¹⁰. Say that they did produce sequences which passed all the requisite tests for randomness. Now say that an agent for a rival company interfered with the set-up by adding an extraneous signal to the process which caused the product to now fail some of the tougher tests for randomness.

That agent evidently introduced some noise into the system, but it was not purely random because it *decreased* the randomness of the result. Here randomness and noise are not the same but are opposed.

What is the case is that sometimes (as in the electrical engineering case above) it is useful to think of this unmodelled residual as arbitrary (particularly if one has good reason to believe that it is irrelevant to the modelling task in hand) and can be usefully represented by a random distribution. This is a natural thing to do since randomness is a positive way to think about noise which otherwise is a negative (what we don't want/understand/represent). In many modelling techniques (e.g. many simulations), it is useful to use a random source to 'stand-in' for parts of the model thought to introduce such arbitrariness. Then one can use techniques such as 'Monte Carlo Sampling' to try and separate out the effects of these parts from the tendencies exhibited by the rest of the model.

However arbitrariness and randomness are not the same as one can not guarantee that an arbitrary signal will not suddenly exhibit some pattern, which may turn out to be significant. This is shown clearly in the example of the 'Millennium Bridge' in London.

Example 4: The Millennium Bridge

The Millennium Bridge is a beautifully elegant structure that spans the river Thames from the bank below St. Paul's to the Tate Modern gallery. This was carefully designed and the design was extensively simulated before being built. However, a few weeks after it was opened it had to be closed again for dampeners to be fitted, as it was prone to oscillations that were deemed a danger to the public. It seems that what happened is that in the simulations of the design the movements of people were assumed to be random, but with large numbers of people they reacted to small movements in the bridge so as to amplify the movement. The small oscillations had the effect of synchronising peoples' reactions and hence having a much greater effect. The result was that the unmodelled reactions of individuals and their coupling via small swayings of the bridge were far from random. Indeed, if they had not happened to cause this particular effect they could have safely regarded as arbitrary and hence modelled using a random 'stand-in', however that did not turn out to be the case (ARUP 2001, BBC News 2000).

¹⁰ There are such companies – I read of a Californian firm which claimed to use light detectors as a source of noise to provide the seeds for algorithms that produce 'high-quality' random numbers for sale (Peterson 1997).

6 Noise and Context

Let us look back at the picture of noise that we started with, that of electrical noise originating in electronic circuits. In one sense this is a very odd usage – noise here does not come from an arbitrary source but is part of the intrinsic nature of the electrical components. From a more objective point of view it is amazing that it is possible to construct ensembles of components that act so as to effect the desired transformations with such an astounding degree of accuracy. It would be less strange if such an ensemble only produced white noise whatever its input than it transforms the inputs to the desired outputs with such accuracy. It is only due to the huge scientific and engineering effort that has led to a situation where we can expect such perfection and hence reify any shortcomings as a separate entity called ‘noise’.

Similar shifts occur in other situations; if one is trying to listen to person A talking then person B talking may be noise, and vice versa. What is noise depends on who one is trying to listen to.

A possible explanation for this is the context-dependency of modeling/ understanding. Any model of observed phenomena has a set of conditions under which it is effectively applicable. Not all of these are explicitly included in models but rather the kind of context where a model is applicable is recognisable. The context ensures that, on the whole, these implicit conditions hold and that the model can be represented in a relatively simple and manageable form. The ‘fuzziness’¹¹ of the context recognition allows the model content to be relatively well-defined and ‘crisp’. This crispness allows us to reason about the content of such models. This necessary context-dependency in modeling is quite a separate matter from whether the model is generally applicable *in theory*. For example, the laws of Newtonian physics *presumably* hold to an astounding degree of accuracy (in circumstances where relativistic and quantum effects are negligible), but may not be practically *applicable* in a situation where well-defined objects are difficult to identify.

The context-dependency of learning and application of knowledge and models makes these processes much more feasible. However, this ‘trick’ only works when the conditions under which a model is learnt are effectively recognisable, so that one knows when one can reliably apply a model. For more on the pragmatic roots of context see (Edmonds 1999b).

Error in a model's output's compared with observations could be due to several things: it could be due to sheer model error within the context; the context itself could have been wrongly identified; or the error could originate from without the assumed context. A coherent picture of noise comes about when the source originates from outside the assumed context – we do not want to reject the model (this still works at some level), and the assumed context seems to be the appropriate one (many other models associated with this context are working well) so the source of the error must lie elsewhere.

This picture of an extra-contextual cause of model error accounts for all the properties of noise we have identified above. In turn it accounts for:

¹¹ I am using the term ‘fuzzines’ here in a loose way, and not necessarily in the sense of Zahdeh (1965).

- The arbitrariness of noise, since it comes from outside the context (since a modelling context is supposed to contain everything pertinent);
- The dependency of the identification of noise on context – thus in the “Two person’s talking” example for each person the context focused on excluded the other talking;
- That the noise has to be able to interfere in the foreground model, since otherwise it has not entered the context as in the “Car Engine” example it depended on whether we imagined the sound escaping into our context (unless we imagined it doing so);
- That the shift from detecting the random fluctuations in the potential difference across a piece of conducting material at thermal equilibrium to the attempt to reduce interference with an audio signal in an amplifier corresponds to the reification of noise.

7 Noise and the “Un-Excluded Middle” between Randomness and Determinism

The thesis of determinism is widely held. To a thorough determinist any randomness is merely a result of incomplete modelling. In other words, *all* apparent randomness is *only* an unmodelled residue¹² – in principle it could all be modelled. Quantum mechanics forces many to accept that there is also irreducible randomness in the universe¹³. Thus the assumption of many is that, in principal everything can be satisfactorily modelled as deterministic processes, except for a random residual – that is the world is neatly divided between the deterministic and the *purely* random. Certainly there are many that assert that almost all *macroscopic* events and processes are essentially deterministic¹⁴, in the sense that all relevant parts of the process can be satisfactorily represented as a deterministic process and the rest is not significant and so representable by random noise.

This picture coincides with the common modelling practice of using a pseudo-random source as a stand-in for an unmodelled residual. This can be an appropriate approach when there is good reason to suppose that the residual *is* arbitrary to the modelled process so that the pattern of this residual does not make any significant difference to the model outcomes (which depends on the model use). A conflation of this modelling approach and a deterministic bias seems to lead some to the conviction that *any* unmodellable residual (a residual that is not capturable in a model in theory) *must be* random.

A particular case of such a conviction is that anything that is not deterministic *must be* random. That is the view that the world is divided between the random and the determined. In this view it may be that some things that appear random turn out to be

¹² Indeed Compagner (1991) argues in the reverse direction – namely that randomness can be defined as what is uncorrelated.

¹³ Of course this does not convince convinced die-hard determinists who live in hope that quantum mechanics will be reduced to deterministic principles someday, as with Einstein’s famous quip “God does not play dice with the universe”.

¹⁴ Despite the evidence from chaotic processes and models.

merely very complex (i.e. we find a way to explicitly model the phenomena where we previously used a random/statistical proxy), and it may be that some things that appear deterministic are so only because of some grand “averaging out” of randomness at the detailed level (such as the atoms in a gas), but ultimately that is all there is: deterministic processes and random ones.

Such a view is closely related to the view that there are no essentially correct models that are necessarily context-dependent. That is to say, that it is always possible to reformulate a model to arbitrarily widen its scope – so that an essentially context-free model is possible. Given such an assumption it is always possible to claim that any unmodelled residual can be eliminated by expanding the model. A slightly weaker form of this assumption is that any *non-random* unmodelled residual is thus eliminable.

However if, on the other hand, some models are intrinsically context-dependent then this is far from necessary, the residual might not be modellable from within any appropriate context and the model might not be possible from within other (e.g. wider) contexts. Another way of putting this is that even the most appropriate context is not hermetically sealed, for there is inevitably some ‘interference from without. In this case the ‘interference’ might not be modellable but there is no reason to suppose that it *is best* ‘represented’ as a random term either¹⁵.

In many mundane cases it is clear that ‘noise’ does not have to be random. The noise of a bus in the street that is interfering with listening to a radio programme is not random just as the “ringing” of transistors after they have switched is not random. Neither are many chaotic systems random, as one can tell by an inspection of their mechanism, even if it is difficult to distinguish their outcomes from random outcomes¹⁶. As (Boffetta *et al* 2002) puts it: “the distinction between chaos and noise ... makes sense only in very peculiar cases, e.g., very low-dimensional systems”.

One response of those who model electrical circuits to the inadequacy of modelling all electrical noise as “white noise” (that is noise which includes, at random, all frequencies) is to enrich the modelling of noise with properties other than the pure form of regulated randomness that characterises “white noise”. As Ritter (2003) put it:

“Both are fundamental sources of “white” noise, meaning that we have a deep statistical understanding of how these sources behave. Unfortunately, this may not be particularly useful if the noise we have is actually due to variable processing-related problem”.

Thus a variety of kinds of noise have now been developed, including “grey” and “black”. If it is right that noise is inherently context-dependent and not necessarily random then there will be no “new” kind of proxy for noise that will *always* be applicable.

¹⁵ Presumably the reason our cognition has divided the world into such contexts means that the amount of critical inter-contextual interference is minimal for its purposes, however the purposes of scientific modelling may be very different and this sort of pragmatic division of the world is heuristic in nature – it is not something that can be relied upon.

¹⁶ This distinction between determining randomness via inspection of the mechanisms vs. by testing the outcomes is made clear in (L’Ecuyer 1992). This is interesting because it is precisely the aim of writing a random-number generating algorithm to separate the external and internal contexts w.r.t. to the generator – from the outside it looks random, whilst from the inside it is deterministic.

Determining what might be the most appropriate proxy for an unmodelled residual is an extremely difficult problem, one where it is difficult, *in principle*, to know what is most appropriate because this would require knowing something about modelling the unmodelled residual, which is not easily amenable to explicit modelling.

8 Conclusion

Noise, in its more general usage can be usefully thought of as extra-contextual interference in what is being modelled. All effective modelling is context-dependent, and no context is completely “water tight” to influences from outside the context, influences that – due to the nature of context – are unmodellable. It is these influences that may interfere with the accuracy (and, indeed, applicability) of our model – we call these influences “noise”. We often model these influences with a random “proxy” because randomness is also unmodellable (in the sense of point-by-point detail).

Although many take randomness to be an essential property of noise, it should be clear from the analysis above that noise is not necessarily *random* but can be merely *arbitrary* with respect to the target. The confusion comes about because we use often randomness as an *archetypal model* for noise. Noise may be of almost any nature considered in isolation – it is only the poverty of our imagination that insists on casting it as its archetype.

References

1. Anderson, P.W.: Some Thoughts About Distribution in Economics. In: Arthur, W.B., Durlaf, S.N., Lane, D.A. (eds.) *The Economy as an Evolving Complex System II*, p. 566. Addison-Wesley, Reading (1997)
2. Andersen, J.V., Sornette, D.: Have Your Cake and Eat It, Too: Increasing Returns While Lowering Large Risks! *Journal of Risk Finance* 2(3), 70–82 (2001)
3. ARUP, The Millennium Bridge. ARUP’s (the engineer’s) site on the Millennium Bridge (2001) (accessed May 15, 2003), <http://www.arup.com/millenniumbridge/>
4. BBC News, The Millennium Bridge. A BBC News report on the Millennium Bridge (2000) (accessed May 15, 2003), http://news.bbc.co.uk/hi/english/static/in_depth/uk/2000/millennium_bridge/
5. Black, F.: Noise. *Journal of Finance* 41, 529–543 (1986)
6. Blake, W.J., Kaern, M., Cantor, C.R., Collins, J.J.: Noise in eukaryotic gene expression. *Nature* 422(6932), 633–637 (2003)
7. Boffetta, G., Cencini, M., Falcioni, M., Vulpiani, A.: Predictability: a way to characterize complexity. *Physics Reports* 356(6), 367–474 (2002)
8. Cartwright, N.: *How the Laws of Physics Lie*. Oxford University Press, Oxford (1983)
9. Compagner, A.: Definitions of Randomness. *American Journal of Physics* 59(8), 700–705 (1991)
10. DeLong, J.B., Shleifer, A., Summers, L., Waldmann, R.J.: Noise trader risk in financial markets. *Journal of Political Economy* 98, 703–738 (1990)
11. Dichev, I.: News or noise? Estimating the noise in the US News university rankings. *Research In Higher Education* 42(3), 237–266 (2001)

12. Edmonds, B.: Modelling Bounded Rationality In Agent-Based Simulations using the Evolution of Mental Models. In: Brenner, T. (ed.) *Computational Techniques for Modelling Learning in Economics*, pp. 305–332. Kluwer, Dordrecht (1999a)
13. Edmonds, B.: The Pragmatic Roots of Context. In: Bouquet, P., Serafini, L., Brézillon, P., Benercetti, M., Castellani, F. (eds.) *CONTEXT 1999*. LNCS, vol. 1688, pp. 119–132. Springer, Heidelberg (1999b)
14. Einstein, A.: *Investigations on the Theory of Brownian Movement*. *Annalen der Physik* 17, 549 (1905) (A.D. Cowper transl.), Methuen, London (1926) and Dover, New York (1956)
15. Gupta, M. (ed.): *Electrical Noise: Fundamentals and Sources*, pp. 51–58. IEEE Press, Los Alamitos (1977)
16. Habermas, J.: *Theorie und Praxis*. Suhrkamp Verlag, Frankfurt am Main (1963) (English trans. Viertel, J.: *Theory and practice*. Polity Press (1988))
17. Hesse, M.B.: *Models and Analogies in Science*. London, Sheed and Ward (1963)
18. Hughes, R.G.: Models and Representation. *Philosophy of Science* 64(proc.), S325–S336 (1997)
19. Kaneko, K.: Globally Coupled Chaos Violates the Law of Large Numbers but not the Central Limit Theorem. *Physics Review Letters* 65, 1391–1394 (1990)
20. Kolmogorov, A.: Three Approaches to the Quantitative Definition of Information. *Problems of Information Transmission* 1, 1–17 (1965)
21. L'Ecuyer, P.: Testing Random Number Generators. In: *Proceedings of the 1992 Winter Simulation Conference*, pp. 305–313 (1992)
22. Peterson, I.: Lava lamp randomness. *Science News* 152, 92 (1997), <http://www.sciencenews.org/20010512/mathtrek.asp>
23. Pierce, J.: Physical Sources of Noise. In: *Proc. IRE*, vol. 44, pp. 601–608 (1956) (reprinted in (Gupta 1977), 51–58)
24. Ritter, T.: Random Electrical Noise: A Literature Survey (2003) (accessed May 15, 2003), <http://www.ciphersbyritter.com/RES/NOISE.HTM>
25. Suppes, P.: Models of Data. In: Nagel, E., Suppes, P., Tarski, A. (eds.) *Proceedings of the 1960 International Congress on Logic Methodology and the Philosophy of Science*, pp. 252–261. Stanford University Press, CA (1962)
26. van Kampen, N.G.: Foreword: The Theory Of Noise. *Fluctuation and Noise Letters* 1(1), 3–7 (2001)
27. Wong, H.: Low-frequency noise study in electron devices: review and update. *Microelectronics Reliability* 43(4), 585–599 (2003)
28. Zadeh, L.H.: Fuzzy Sets. *Information and Control* 8, 338–353 (1965)

Author Index

Ahrweiler, Petra	12	Moss, Scott	48
Conte, Rosaria	29	Neumann, Martin	69
David, Nuno	117	Ormerod, Paul	130
Doran, Jim	153	Polhill, J. Gary	99
Edmonds, Bruce	169	Rosewell, Bridget	130
Frank, Ulrich	1	Roth, Camille	89
Gilbert, Nigel	12	Schmid, Alex	60
Gotts, Nicholas M.	99	Squazzoni, Flaminio	1
Lorenz, Tobias	141	Troitzsch, Klaus G.	1