Studies in Applied Philosophy, Epistemology and Rational Ethics

SAPERE

Emiliano Ippoliti
Fabio Sterpetti
Thomas Nickles *Editors*

Models and Inferences in Science

Springer

Studies in Applied Philosophy, Epistemology and Rational Ethics

Volume 25

About this Series

Studies in Applied Philosophy, Epistemology and Rational Ethics (SAPERE) publishes new developments and advances in all the fields of philosophy, epistemology, and ethics, bringing them together with a cluster of scientific disciplines and technological outcomes: from computer science to life sciences, from economics, law, and education to engineering, logic, and mathematics, from medicine to physics, human sciences, and politics. It aims at covering all the challenging philosophical and ethical themes of contemporary society, making them appropriately applicable to contemporary theoretical, methodological, and practical problems, impasses, controversies, and conflicts. The series includes monographs, lecture notes, selected contributions from specialized conferences and workshops as well as selected Ph.D. theses.

Advisory Board

More information about this series at http://www.springer.com/series/10087

Emiliano Ippoliti · Fabio Sterpetti
Thomas Nickles
Editors

Models and Inferences in Science

Springer

Editors
Emiliano Ippoliti
Dipartimento di Filosofia
Sapienza University of Rome
Rome
Italy

Thomas Nickles
Department of Philosophy
University of Nevada
Reno, NV
USA

Fabio Sterpetti
Dipartimento di Filosofia
Sapienza University of Rome
Rome
Italy

ISSN 2192-6255 ISSN 2192-6263 (electronic)
Studies in Applied Philosophy, Epistemology and Rational Ethics
ISBN 978-3-319-28162-9 ISBN 978-3-319-28163-6 (eBook)
DOI 10.1007/978-3-319-28163-6

Library of Congress Control Number: 2015959591

Printed on acid-free paper

This Springer imprint is published by SpringerNature
The registered company is Springer International Publishing AG Switzerland

Contents

Modeling and Inferring in Science

Emiliano Ippoliti, Thomas Nickles and Fabio Sterpetti

Science continually contributes new models and rethinks old ones. The way inferences are made is constantly being re-evaluated. The practice and achievements of science are both shaped by this process, so it is important to understand how models and inferences are made.

Despite the relevance of models and inference in scientific practice, these concepts are not only multifaceted but also in some sense their definition, role and purpose still remain controversial in many respects.

Let us start with the notion of model. Frigg and Hartmann, for instance, state that:

> Models can perform two fundamentally different representational functions. On the one hand, a model can be a representation of a selected part of the world (the 'target system'). […]. On the other hand, a model can represent a theory in the sense that it interprets the laws and axioms of that theory. These two notions are not mutually exclusive as scientific models can be representations in both senses at the same time.[1]

It seems that the concept of 'model' is so wide that it cannot be grasped by means of a single, clear definition, and thus its meaning is still controversial. In effect, there are several definitions of what models are, which often sharply diverge (see e.g. Krause and Bueno 2007, p. 187). For example, Bailer-Jones states that a "model is an interpretative description of a phenomenon that facilitates access to that phenomenon," and that interpretative descriptions may rely "on idealizations or simplifications or on analogies to interpretative descriptions of other phenomena." Moreover, models can "range from being objects, such as a toy airplane, to being

[1]Frigg and Hartmann (2012), § 1.

E. Ippoliti (✉) · F. Sterpetti
Sapienza University of Rome, Rome, Italy
e-mail: emi.ippoliti@gmail.com

F. Sterpetti
e-mail: fabio.sterpetti@uniroma1.it

T. Nickles
University of Nevada, Reno, NV, USA
e-mail: nickles@unr.edu

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_1

theoretical, abstract entities, such as the Standard Model of the structure of matter and its fundamental particles" (Bailer-Jones 2009, pp. 1–2). Along this line, models may be easily conceived as instruments, 'neither true nor false': instead, they are useful heuristic devices, which often are effective even when they are 'false'.

On the contrary, it has been argued that models cannot be interpreted as useful heuristic devices, because if "theories are vehicles of scientific knowledge, then so too must models be" (Suppe 2000, p. S109). The reason for such a claim is that if knowledge is intended as being related to the truth, and theories are vehicles of knowledge, models of such theories have to be true, nor just metaphor-like or heuristic devices. In fact, in model theory models of a theory *make true* the axioms of such theory. Since those who adopt the semantic view of theories adopt the model theoretic concept of 'model', they cannot think of models as heuristic devices (Morrison 2009). For example, Suppes claims that "the concept of model in the sense of Tarski may be used without distortion and as a fundamental concept" in scientific and mathematical disciplines, and that "the meaning of the concept of model is the same in mathematics and the empirical sciences" (Suppes 1961, p. 165). According to Suppe, "Suppes' claim is that the Tarski concept of a model is a common formal framework for analysis of various uses of models in science and mathematics" (Suppe 2000, p. S111).

Many authors have criticized this conflation of different senses attached to the term 'model' (Thomson-Jones 2006). But the problem is that if we decouple the concept of model used in model theory from that used for heuristic purposes in scientific practice, then it is difficult to maintain some of the traditional realist claims about the truth of our best scientific theories that many philosophers of science subscribe to. Indeed, the best tools to describe the idea that our best theories correctly 'correspond' to the world have been for a long time Tarski's theory and the notion of 'isomorphism' (da Costa and French 2003).

In effect, the move of denying the identity of the concept of model used in mathematics and that used in scientific practice, by trying to develop a more subject- and context-dependent notion of model centered on the notion of 'representation' instead of that on that of 'isomorphism', has faced two main objections.

First, there is the argument from realist-minded philosophers that such a motive implies or at least invites a sort of instrumentalism that is not able to preserve the objectivity of science, and thus risks to open the door to skepticism or relativism. Second, again coming from some realist philosophers, is that the notion of representation used by the instrumentalists may be in its turn accounted for in terms of isomorphism, and so that the notions of model used in mathematics and in scientific practice are not really distinct and may be in the ultimate analysis reduced to one (French 2003).

As concerns the notion of inference, its role, nature and purpose are at stake as well, since the orthodox viewpoint put forward by the analytic tradition, modeled on mathematical logic, displayed more and more weaknesses, especially in the attempt to account for the growth of knowledge (see e.g. Cellucci 2013; Ippoliti 2014). For an increasing number of philosophers, this problem requires a completely new approach to the concepts of knowledge and inference, both internally and externally.

More specifically, the standard view of the notion of inference is "formulated by Hintikka and Sandu as follows: 'Inferences can be either deductive, that is, necessarily truth preserving, or ampliative, that is, not necessarily truth preserving'" (Cellucci 2013, p. 295). Such a distinction is *internally* inadequate, since it does not produce a cogent classification of the various kinds of inference. In particular it does not work for abduction. In fact, abduction, as many people understand it, is neither ampliative nor truth preserving, and hence it is a counter-example to this standard way of conceiving inferences. If we accept the orthodox distinction between deductive rules (truth preserving) and ampliative rules (non-deductive, and hence not truth preserving), it turns out that abduction "belongs to a different category because, on the one hand, like deductive rules, it is non-ampliative, but, on the other hand, unlike them, it is not truth preserving" (Ibidem, p. 302).

On the other side, the standard view is unsatisfactory also *externally*, that is, with respect to the nature, role and purposes of knowledge. For, in the end, it does not account for the growth and ampliation of knowledge. Mathematical logic, the model of the analytic tradition, is a tool explicitly designed to systematize and justify what is already known. It does not aim at producing genuinely new knowledge, since its purpose is to provide a secure foundation for our scientific knowledge, in particular mathematics; and the method to do that is the deductive method.

First, mathematical logic fails as a means of justification, in virtue of a careful reading of the limitative results in general, and of the consequences of Gödel's incompleteness theorems in particular (see Cellucci 2013).

Second, the analytic tradition and mathematical logic essentially draw on a restriction on the scope of logic, and hence inference, with respect to Plato, Aristotle, Descartes and Kant, which turned out to be detrimental to its role in the scientific research.

A promising way out to these difficulties is to approach the notion of inference using different notions, namely the one of containment instead of truth preservation and the one of vindication instead of validation. We will sketch here the former.

We can produce a more cogent classification of inferential rules in terms of ampliative and non-ampliative ones. The former, like induction or analogy, are such that their conclusions are not *contained* in the premises, the information in the conclusion goes beyond the information in the premises. And because of that they can go wrong, even if they have heuristic power. The latter, like the deductive rules, are such that the conclusion is contained in the premises, that is, the conclusion either is literally a part of the premises, or entails nothing that is not already entailed by the premises. For instance, in *Modus Ponens* the conclusion B is literally included in the premises A, and A $\rightarrow$ B. Therefore, deductive rules, as non-ampliative rules, have no heuristic power. For a new idea (B in this case) must already be available before the inference can be constructed. It is not, therefore, an inference *to* B as new knowledge. It goes without saying that it does not mean that they are useless. As a matter of fact, since the conclusion of a deductive rule makes explicit all or part of what is contained in the premises, it enables us to establishing

that its conclusion is plausible, facilitating the comparison of the premises with experience.

The bottom line here is that there is no consensus on how models and inferences are to be understood. Thus, inquiring into the nature and role of models and inferences is at the top of the philosophical agenda, and tellingly several works have been devoted to this issue in recent years (Humphreys and Imbert 2012; Meheus and Nickles 2009; Suarez 2009; Morgan and Morrison 1999).

So the attempt to understand the ways models and inferences are made basically opens two roads. The first one is to produce an analysis of the role that models and inferences play in science—how sciences use models and inferences to inquire into the world. The second one is to produce an analysis of the way models and inferences are constructed—how to model the way that scientific knowledge is pursued, especially in the light of what science tells us about our cognitive abilities.

This volume goes both ways. In the exploration of the nature and role of models and inferences, the contributed papers focus on different aspects of both the way in which models and inferences are relevant to science and the way in which science is relevant to rethinking what models and inferences are, and how models and inferences are made. In fact, the collected papers deal with issues such as: the role of the models in scientific practice; how science shapes our conceptions of models; how to model the pursuit of scientific knowledge; the relation between our conception of models and our conception of science; models and scientific explanation; models in the semantic view of theories; the applicability of mathematical models to the world; the relation between models and inferences; models as a means for acquiring new knowledge.

In dealing with those issues, the collected papers clearly underline that in order to better understand what models are it is crucial to investigate how our accounts of models and inferences are related to the way in which we analyse human knowledge, and specifically scientific knowledge.

Knowledge is indeed a crucial issue when dealing with models and inferences. To see this point let us consider some well-known and debated issues in philosophy of science.

The discussion over the nature of abduction, and the related ‘Inference to the Best Explanation’, that has taken place in recent years (Magnani 2009; Aliseda 2006; Lipton 2004) can be seen as an example of the relevance of the way in which inferences are analysed for the way in which science is characterized, and the reciprocal relevance of the view about science that we adopt for the definition of our ideas with regard to the nature of inferences. Whether abduction has to be considered an ampliative inference, and whether abductive reasoning has to be considered an acceptable form of scientific reasoning, are questions deeply related to the dispute over scientific realism, i.e., the way in which scientific knowledge has to be understood. Different ways of conceiving the same inference are due to the different conception of knowledge that one can deploy. And the concept of knowledge that one can accept is at its turn related to the way in which one conceives of the nature and the role of certain inferences.

Another example of the connection between the way in which inferences are characterized and the way in which science is analysed is the issue of the ampliativity of deduction (Prawitz 2014). To take a stance on that issue clearly makes a great difference for the way in which one conceives the scientific method. In fact, if deduction may give us new knowledge, i.e. something more than what is already contained in the premises, then the method of science may be more easily conceived of in axiomatic-deductivist terms. If, on the contrary, deduction is considered not to be ampliative, then an axiomatic-deductivist view cannot account for the process of knowledge ampliation. And subscribing to a specific view on how the scientific method has to be characterized has a great relevance for our considering deduction as ampliative or not, and thus has a great relevance on the way in which knowledge is intended (Cellucci 2013).

But even science, i.e., our recent scientific acquisitions, is relevant to logic and the way in which we conceive of the nature of inferences. For example, naturalism seems to be a mainstream tendency in contemporary philosophy, but the impact that a naturalistic stance on logic, inspired by recent work on human cognitive structures and evolution, could have on the way in which logic is conceived of is not yet clear (Schechter 2013; Dutilh Novaes 2012; Pelletier et al. 2008).

There is a similar relation between the way in which we conceive of mathematics and science. For example, as we have already seen above, despite the wide acceptance of the semantic view of theories, which, roughly speaking, says that a theory is the class of its models, the difficulties of making such a definition compatible with the conception of model usually accepted in model theory have not been overcome (Halvorson 2012). Moreover, models are normally understood by the semanticists as mathematical models. Thus, the problem of the relation between a theory and the world is connected to the issue of the relation between mathematics and the world. This means that the question about the role of models in science is ultimately related to the question of the nature of the relation between mathematics and the world, and thus to the question about the nature of mathematics (Cellucci 2013).

This suggests that, as in the case of logic, science not only *uses* mathematics, but even puts pressure on philosophers to rethink what mathematics is, so to make our conception of what mathematics is more compatible with what science tells us about the way the world is. And doing so, in turn, can even lead us to rethink what science is. Thus, not only our models and inferences, but also our way of modelling our models and inferences are worth being continuously investigated.

The papers collected in this volume are devoted precisely to the task of rethinking and better understanding what models and inferences are. It will be useful to describe their content in some detail.

Sorin Bangu's paper, *On 'The Unreasonable Effectiveness of Mathematics in the Natural Sciences'*, deals with Eugene Wigner's famous claim that the appropriateness of the language of mathematics for the formulation of the laws of physics is a miracle (Wigner 1960). Bangu reconstructs Wigner's argument for the unreasonable effectiveness of mathematics and takes into account six objections to its

soundness. After having shown that those six objections are weaker than it is usually thought, he raises a new objection to Wigner.

Thomas Nickles, in his *Fast and Frugal Heuristics at Research Frontiers*, investigates how we should model scientific decision-making at the frontiers of research. Nickles explores the applicability of Gigerenzer's 'fast and frugal' heuristics to the context of discovery. Such heuristics require only one or a very few steps to a decision and only a little information. While Gigerenzer's approach seems promising in accounting for the context of discovery, given the limited resources available in frontier contexts, it nevertheless raises challenging questions, since it seems that, according to this view of frontier epistemology, we find ourselves in the quite paradoxical situation in which the way forward may be to make sparse information even sparser.

Fabio Sterpetti's *Scientific Realism, the Semantic View and Evolutionary Biology* deals with the difficulties which arise when we try to apply structural realism and the semantic view of theories to some philosophical issues peculiarly related to biology. Given the central role that models have in the semantic view, and the relevance that mathematics has in the definition of the concept of model, Sterpetti focuses on population genetics, which is one of the most mathematized areas in biology, to assess French's proposal (French 2014) of adopting structural realism in dealing with biology.

Emily Grosholz's *Models of the Skies* examines the development of models of astronomical systems, beginning with the early 17th century models of the solar system, and ending with late 20th century models of galaxies. More precisely, models by Kepler, Newton, Laplace, Clausius, Herschel, Rosse, Hubble, Zwicky, and Rubin are taken into account. In each case she emphasizes the distinction and the interaction between the aims of reference and analysis, and the ways in which disparate modes of representation combine to enlarge scientific knowledge.

Carlo Cellucci, in his *Models of Science and Models in Science*, deals with the issue of how it is possible to model science. Indeed, with regard to science, one may speak of models in two different senses, i.e. 'models of science' and 'models in science'. A model of science is a representation of how scientists build their theories, a model in science is a representation of empirical objects, phenomena, or processes. Cellucci considers five models of science: the analytic-synthetic model, the deductive model, the abstract deductive model, the semantic model, and the analytic model. After presenting them, he assesses to what extent each of them is capable of accounting for models in science.

Raffaella Campaner's *Mechanistic Models and Modeling Disorders* deals with the debate on how disorders should be modeled, and focuses on some issues arising from modeling neuropsychiatric disorders. More precisely, she discusses some models of attention deficit hyperactivity disorder (ADHD). The main aspects of such models are analyzed in the light of the philosophical debate about mechanistic models. The paper highlights how the neo-mechanist accounts of models can only partly capture the many aspects entering the dynamics of modeling disorders in an actual medical scenario.

Sergio Caprara's and Angelo Vulpiani's paper, *Chaos and Stochastic Models in Physics*, deals with the issue of clarifying the distinction between determinism and predictability. In order to show that the two concepts are completely unrelated, Caprara and Vulpiani analyse the Lyapunov exponents and the Kolmogorov-Sinai entropy and show how deterministic chaos, although it possesses an epistemic character, is not subjective at all. They also show how this is useful to shed light on the role of stochastic models in the description of the physical world.

Emiliano Ippoliti's paper, *Ways of Advancing Knowledge. A Lesson from Knot Theory and Topology*, investigates the ways of advancing knowledge focusing on the construction of several approaches put forward to solve problems in topology and knot theory. More precisely, Ippoliti considers two problems: the classification of knots and the classification of 3-manifolds. Examining the attempts made to solve those problems, Ippoliti is able to specify some key features of the ampliation of knowledge, such as the role of representation, theorem-proving and analogy, and to derive some considerations on the very nature of mathematical objects.

Juha Saatsi's paper, *Models*, *Idealisations*, *and Realism*, deals with the difficulties that, for the scientific realist, derive from the role that idealizations and abstractions play in models. Indeed, realists maintain that predictively successful models tell us the truth about the unobservable world. But how should the realist construe the way in which models latch onto unobservable reality? This is a problem, since models essentially incorporate various kinds of idealisations and approximations that cannot be interpreted realistically and that are indispensable to both their predictive and their explanatory use. Saatsi tries to face such a challenge by arguing that it is the modal character of idealisations that accounts for their utility from a realist perspective.

In *Modelling Non-Empirical Confirmation* Richard Dawid argues that non-empirical theory confirmation plays an important role in the scientific process and that it should be considered an extension of empirical confirmation. Since confirmation is mostly understood in Bayesian terms, Dawid proposes a formalization of non-empirical confirmation within a Bayesian framework that demonstrates that non-empirical confirmation does have the same structural characteristics of empirical theory confirmation. The No Alternative Argument (Dawid et al. 2015) is then illustrated and debated.

Reuben Hersh's paper *Mathematics as an Empirical Phenomenon*, *Subject to Modeling* deals with the issue of modeling mathematics. Indeed, philosophy of mathematics deals with models of mathematics, which is in large part already a model, because much of mathematics is a model of physical action. Arithmetic, for instance, models the human action of counting. Hersh's suggestion is that in order to facilitate the creation of a unified field of inquiry on mathematics, philosophers should start thinking of their work as model-building instead of arguing for their chosen position against opposing positions.

Lorenzo Magnani's paper, *Scientific Models Are Distributed and Never Abstract: A Naturalistic Perspective*, analyses several definitions of models: from the classical ones, which see models as abstract entities and idealizations, to the more recent, which see models as fictions, surrogates, credible worlds, missing

systems, make-believe, parables, epistemic actions. Magnani reveals some of their epistemological inadequacies, sometimes by appealing to recent results in cognitive science. Magnani specifically addresses epistemological relying on recent results on the role of distributed and abductive cognition.

Kahindo Kamau's and Emily Grosholz's paper *The Use of Models in Petroleum and Natural Gas Engineering* inquires how adequate are some of the fundamental models in the science of petroleum and natural gas engineering. The authors try to unveil what assumptions were made as the models were created. They claim that a good account of the adequacy of models must be strongly pragmatist, for the questions related to their adequacy cannot be answered properly without paying attention to human purposes. They also claim that many of the distortions and over-simplifications in these models are in fact intentional and useful, when we examine the models in the light of their pragmatic aims.

References

Aliseda, A.: Abductive Reasoning. Springer, Dordrecht (2006)
Bailer-Jones, D.M.: Scientific Models in Philosophy of Science. University of Pittsburgh Press, Pittsburgh (2009)
Cellucci, C.: Rethinking Logic. Logic in Relation to Mathematics, Evolution, and Method. Springer, Dordrecht (2013)
da Costa, N.C.A., French, S.: Science and Partial Truth. A Unitary Approach to Models and Scientific Reasoning. Oxford University Press, Oxford (2003)
Dawid, R., Hartmann, S., Sprenger, J.: The no alternatives argument. Br. J. Philos. Sci. **66**, 213–234 (2015)
Dutilh Novaes, C.: Formal Languages in Logic. A Philosophical and Cognitive Analysis. Cambridge University Press, Cambridge (2012)
French, S.: A model-theoretic account of representation (or, I don't know much about art… but I know it involves isomorphism). Philos. Sci. **70**, 1472–1483 (2003)
French, S.: The Structure of the World. Metaphysics and Representation. Oxford University Press, Oxford (2014)
Frigg, R., Hartmann, S.: Models in science. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy (Fall 2012 Edition), http://plato.stanford.edu/archives/fall2012/entries/models-science/ (2012)
Halvorson, H.: What scientific theories could not be. Philos. Sci. **79**, 183–206 (2012)
Humphreys, P., Imbert, C. (eds.): Models, Simulations, and Representations. Routledge, New York (2012)
Ippoliti, E. (ed.): Heuristic Reasoning. Springer, Cham (2014)
Krause, D., Bueno, O.: Scientific theories, models, and the semantic approach. Principia **11**, 187–201 (2007)
Lipton, P.: Inference to the Best Explanation. Routledge, London (2004)
Magnani, L.: Abductive Cognition. The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning. Springer, Berlin (2009)
Meheus, J., Nickles, T. (eds.): Models of Discovery and Creativity. Springer, Dordrecht (2009)
Morgan, M.S., Morrison, M. (eds.): Models as Mediators. Perspectives on Natural and Social Science. Cambridge University Press, Cambridge (1999)

Morrison, M.: Fictions, representations, and reality. In: Suárez, M. (ed.) Fictions in Science. Philosophical Essays on Modeling and Idealization, pp. 110–135. Routledge, New York (2009)

Pelletier, F.J., Elio, R., Hanson, P.: Is logic all in our heads? From naturalism to psychologism. Stud. Logica. **88**, 3–66 (2008)

Prawitz, D.: The status of mathematical knowledge. In: Ippoliti, E., Cozzo, C. (eds.) From a Heuristic Point of View. Essays in Honour of Carlo Cellucci, pp. 73–90. Cambridge Scholars Publishing, Newcastle upon Tyne (2014)

Schechter, J.: Could evolution explain our reliability about logic? In: Gendler, T.S., Hawthorne, J. (eds.) Oxford Studies in Epistemology, vol. 4, pp. 214–239. Oxford University Press, Oxford (2013)

Suarez, M. (ed.): Fictions in Science. Philosophical Essays on Modeling and Idealization. Routledge, New York (2009)

Suppe, F.: Understanding scientific theories: an assessment of developments, 1969–1998. Philos. Sci. **67**, S102–S115 (2000)

Suppes, P.: A comparison of the meaning and use of models in mathematics and the empirical sciences. In: Freudenthal, J. (ed.) The Concept and the Role of the Model in Mathematics and Natural and Social Sciences, pp. 163–177. Reidel, Dordrecht (1961)

Thomson-Jones, M.: Models and the semantic view. Philos. Sci. **73**, 524–535 (2006)

Wigner, E.: The unreasonable effectiveness of mathematics in the natural sciences. Commun. Pure Appl. Math. **13**, 1–14 (1960)

On 'The Unreasonable Effectiveness of Mathematics in the Natural Sciences'

Sorin Bangu

Abstract I present a reconstruction of Eugene Wigner's argument for the claim that mathematics is 'unreasonable effective', together with six objections to its soundness. I show that these objections are weaker than usually thought, and I sketch a new objection.

1 Introduction

In a well-known essay published in 1960, the celebrated physicist Eugene Wigner claimed that "the appropriateness of the language of mathematics for the formulation of the laws of physics" is a "miracle" (Wigner 1960, p. 14). Despite Wigner's immense scientific reputation (he will be awarded the Nobel prize in 1963), the general sentiment is that he hasn't quite succeeded in making a case for the 'miraculousness' of the applicability of mathematics—although everyone agrees that the issue is prima facie intriguing. In fact, the issue was considered so intriguing that several of the brightest minds of theoretical physics (Dirac, Weinberg, Wilczek) found worth engaging with it; moreover, one even gets the impression, upon becoming familiar with the early literature discussing this so-called 'Wigner puzzle', that for a good while after 1960 the conundrum interested more the scientists and the mathematicians than the philosophers. This situation, it seems to me, changed significantly after the year 2000—that is, after the publication, in 1998, of Mark Steiner's landmark book *The Applicability of Mathematics as a Philosophical Problem* (Harvard Univ. Press). Thus, in the last decade or so, partly due to this book's influence, the puzzle has received significantly more attention from philosophers.

And, to be sure, there is no shortage of attempts to (dis)solve the puzzle, fact which accounts for the almost universal skeptical sentiment I mentioned above.

S. Bangu (✉)
University of Bergen, Bergen, Norway
e-mail: sorin.bangu@uib.no

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_2

In this paper, I will end up sharing this sentiment, but not before casting a critical eye on the proposed solutions. I shall proceed as follows. First of all, I'll spell out the puzzle—or, more precisely, a *version* of it as reconstructed from Wigner's essay. The 'unreasonableness' claim will appear as the conclusion of a valid argument, and thus the next natural step will be to inspect the premises. Then, I will identify six different (types of) solutions, each of them attacking one (or more) of these premises. Although these are cogent objections, and certainly raise doubts about the soundness of the argument, they are not decisive; a defender of Wigner's central point will surely feel their force, but will not need to concede defeat. Finally, I will sketch a different (and, as far as I can tell, novel) solution to the puzzle, drawing on what I'll call 'ecological' considerations affecting scientific research.[1]

2 Wigner's Argument

Before we get to discuss Wigner's argument per se, it is important to clarify two aspects of it. First, Wigner talks about the unreasonable effectiveness of 'mathematics' in physics, but what he has in mind is something slightly more specific: the effectiveness of the mathematical *language*—and by this it is pretty clear that he means the effectiveness of a fair amount of *mathematical concepts* (and *structures*), such as complex number, group, Hilbert space, etc. (these are some of his own examples). The second and related question is 'what are these concepts effective *for*?' Wigner's answer is that these concepts are effective for "the formulation of the laws of physics" (1960, p. 6), that is, in *describing* natural phenomena, or, more exactly, certain law-like regularities holding in nature.

This clarification of *what* is effective (mathematical concepts), and what they are effective *for* (describing nature), is necessary in order to distinguish Wigner's concern from a recent proposal of a somewhat similar problem by Steiner (1998). For Steiner, what is primarily (and ultimately mysteriously) effective are mathematical *analogies*, and what they are effective for is the formulation of *novel laws* of physics —that is, laws formulated by analogy with the existent mathematically formulated laws.[2] (The new laws are needed in domains of reality not covered by the existing laws, such as the quantum domain.) Unlike Wigner's, Steiner's main concern is thus the *heuristic* role of mathematics, or its ability to mediate the development of new laws. Here, however, I will put this issue aside, and focus on Wigner alone.[3]

---

[1] I deal with the puzzle in my (2009) and, more thoroughly, in my (2012, Chap. 7). Although there is some overlap between this paper and my treatment of the issue in my book, the current paper offers a different reconstruction of the puzzle. My conclusion, however, is the same—that Wigner's riddle can be (dis)solved.

[2] Grattan-Guiness (2008) seems to me an example, among others, of conflating these separate issues: Wigner's, who focused on the role of mathematics in describing nature, and others' concerns with its role in theory-building.

[3] For my take on Steiner's own argument, see my (2006) and (2012, Chap. 8).

Now, what is Wigner's argument in his 1960 paper? As it happens, this is not immediately clear, since his points are open to a couple of reconstructions. Steiner (1998, pp. 45–6) offered one of the first such careful renderings. He identified two versions of the argument; one is as follows:

> Concepts c_1, c_2, c_3,…, c_n (some listed in the paper; see above for a sample) are unreasonably effective in physics, and these concepts are mathematical.
>
> Hence, mathematical concepts ('mathematics') are (is) unreasonably effective in physics.

This argument is invalid, and if this version is what the critics had in mind then their discontent is understandable. Steiner points out that the conclusion doesn't follow; what follows is a weaker claim, that *some* mathematical concepts are unreasonably effective—and this invites the query as to how this unreasonable effectiveness is related to their being mathematical. Yet, with Steiner, I also believe that a more charitable reconstruction is possible and, taking my cue from his analysis (and also departing from it), I will put one forward below—and call it 'W_A'.

My W_A is meant to be the version of Wigner's concern that fascinated those most brilliant theoretical physicists I named above, and its specificity is that it is a *diachronic*, or *historically-based* reconstruction of his point. I favor this specifically diachronic version since it reflects faithfully the oddity of a certain "situation" (Dirac's word; see below) noticed not only by Wigner, but also by other people, both before and after the publication of his article. Here is what Paul Dirac said in 1939 in his note on 'The relation between mathematics and physics':

> One may describe this situation by saying that the mathematician plays a game in which he himself invents the rules while the physicist plays a game in which the rules are provided by Nature, but as time goes on it becomes increasingly evident that the rules which the mathematician finds interesting are the same as those which Nature has chosen. (1939, p. 124)

This quote is very suggestive, as it encapsulates all the elements I will include in the W_A: the idea that mathematicians 'invent the rules', that what drives this invention is what they find 'interesting' (hence the aesthetic aspects of W_A), and finally the overt reference to the temporal succession. Similar to Dirac's point above, Steven Weinberg writes:

> It is positively spooky how the physicist finds the mathematician has been there before him or her. (1986, p. 725),

where, importantly, what led the mathematicians 'there' was their aesthetical sense:

> [M]athematicians are led by their sense of mathematical beauty to develop formal structures that physicists only later find useful, even where the mathematician had no such goal in mind. […]. Physicists generally find the ability of mathematicians to anticipate the mathematics needed in the theories of physics quite uncanny. It is as if Neil Armstrong in 1969 when he first set foot on the surface of the moon had found in the lunar dust the footsteps of Jules Verne. (1993, p. 125)

Wigner himself talks explicitly in diachronic terms, when referring to physical concepts as discovered

> independently by the physicist and recognized then as having been conceived *before* by the mathematician. (1960, p. 7; my emphasis)

Finally, against this background, this is the W_A:

1. Modern mathematical concepts originate in our (mathematicians') aesthetic preferences.
2. It is unreasonable that these concepts, originating in the subjective aesthetic domain, are effective in the objective domain of physics.
3. And yet this is the case: several physical theories proposed at a later time t' turned out to benefit significantly from the application of mathematical concepts developed at an earlier time t.
4. Therefore, it is unreasonable that modern mathematical concepts (developed up to an earlier time t) are effective in the physics introduced at a later time t'.

Curious as it may seem, such explicit reconstructions of the problem are not common in the literature. It is not always recognized what the Wigner puzzle in fact is, namely a *pre-established harmony* type of mystery[4]: how can it be that such a temporal anticipation of physics by mathematics exists throughout the history of science?

Since the validity of argument W_A is not an issue anymore, the objections have to focus on the truth of the premises. And, as I said, all proposed solutions so far are formulated as attacks on one, or several, of these three premises. I will examine the premises in the next section (and, after that, the solutions).

3 A Closer Look at the Premises

Let us put the premises under a magnifying glass. I will take them in turn. To begin with the first, what does it mean to say that modern mathematics and, more specifically, modern mathematical concepts and structures, have aesthetic origins? That is, what can one make of the claim that modern mathematics is "the science of skillful operations with concepts and rules invented just for this purpose", where the purpose is for mathematicians to "demonstrate [their] ingenuity and sense of formal beauty."? (Wigner 1960, p. 2)[5]

[4]In fact, Bourbaki, when referring to this issue, uses the word 'preadaption'. Here is the entire quote: "Mathematics appears […] as a storehouse of abstract forms—the mathematical structures; and it so happens—without out knowing why—that certain aspects of empirical reality fit themselves into these forms, as if through a kind or pre-adaption." (1950, p. 231) I found this quote in Ginammi (2014, p. 27).

[5]Wigner also writes that mathematical concepts "are defined with a view of permitting ingenious logical operations which appeal to our aesthetic sense … [they are chosen] for their amenability to clever manipulations and to striking, brilliant arguments." (1960, p. 7).

This first premise makes two claims. First, that (i) mathematics is a human *invention*, i.e., the concepts are free inventions of the mind, and also that (ii) among these many free creations, some of them strike the mathematicians as particularly beautiful, interesting, etc.—and thus they are selected, on the basis of these *aesthetic criteria*, to be studied and developed (typically by formulating and proving theorems about them.)

It is important to clarify what 'aesthetic' means in this context. The central idea of an aesthetic preference has to be construed as a rather broad notion. It is an umbrella-term, standing of course for what Wigner himself called above "formal beauty", but also covering a wider gamut of related sentiments such as certain concepts being 'interesting', 'elegant', 'simple', 'deep', 'unifying', 'fruitful', 'stimulating', 'intriguing', etc. Like other important physicists (his friend John von Neumann included; see below), Wigner believes that mathematicians are free to choose what concepts to work with, and they select what they find—in these various guises—'beautiful'.

Thus, to say that the primary creative impulse of a (modern) mathematician is aesthetic is to stress that the concepts and structures she selects to study are

(a) *neither* descriptions of some natural phenomenon,
(b) *nor* tools to help the development of an existing (perhaps incipient) physical theory.

Two examples may clarify the matter here. The invention of *real* analysis (or 'calculus'), by Leibniz and Newton, provides one particularly clear illustration of a mathematical achievement that does *not* have aesthetic origins. On the other hand, the concept *complex number* (and, consequently, *complex* analysis) does qualify as having aesthetical ancestry, since the introduction of complex numbers satisfied clauses (a) and (b) above.[6] The same relation holds in other subfields of mathematics, for instance between Euclidean geometry and its various multi-dimensional generalizations or alternatives. It is also important to understand that this 'mathematical aestheticism' is perfectly compatible with some of the aesthetically-driven mathematicians' *hope* or *desire* that maybe in the future the physicists will find the concepts she studied useful. This kind of attitude (sometimes transpiring in their writings) doesn't make the initial impulse to focus on these concepts and structures less 'pure', i.e., less aesthetical. (We'll get back to this point when we'll discuss the Riemann episode below.)

Returning to the first premise, its two parts have different statuses. Component (i) expresses adherence to a metaphysical view of the nature of mathematics (anti-Platonism), while (ii) sounds more like a factual statement about certain historical/psychological events, or processes: the circumstances of origination, or

[6]Jerome Cardan, who is credited with introducing them in the 16th century, remarked that "So progresses arithmetic subtlety the end of which, as is said, is as refined as is useless." (Cited in Kline 1972, p. 253). According to Kline, neither did Newton regard complex numbers as significant, "most likely because in his day they lacked physical meaning." (1972, p. 254).

invention, of certain concepts. I will leave (i) aside for the moment (I will get back to it in Sect. 4), as it is notoriously difficult to search for justifications for such basic metaphysical commitments—here I'll only focus on (ii). This is a claim that can be vindicated by research into the history of mathematics. This kind of research is available and, as it happens, seems to confirm Wigner. The historian Kline (1972, pp. 1029–31) summarizes the situation as following:

> [G]radually and unwittingly mathematicians began to introduce concepts that had little or no direct physical meaning (…) [M]athematics was progressing beyond concepts suggested by experience (…) [M]athematicians had yet to grasp that their subject … was no longer, if it ever had been, a reading of nature. (…) [A]fter about 1850, the view that mathematics can introduce and deal with rather arbitrary concepts and theories that do not have immediate physical interpretation but may nevertheless be useful, as in the case of quaternions, or satisfy a desire for generality, as in the case of n-dimensional geometry, gained acceptance.[7]

Another way to go about this first premise is to simply ask the (great) mathematicians themselves: do *they* think that a view like Wigner's has any credibility?[8] If the practitioners' avowals are to be given any weight, then aestheticism is supported by quite a few, and prominent mathematicians. Among the most cited such confessions is the one belonging to Richard Hamming (of the 'Hamming code' fame), that "*artistic taste* plays a large role in modern mathematics" (1980, p. 83; author's emphasis)[9]; another belongs to no less a figure than John von Neumann. He makes the point at the end of the paragraph below, worth quoting in full because it also canvasses some important insights into the relation between physics and mathematics. Like Dirac, he talks about a certain "situation":

> The situation in mathematics is entirely different [from physics]. (…) 'Objectively' given, 'important' problems may arise after a subdivision of mathematics has evolved relatively far and if it has bogged down seriously before a difficulty. But even then the mathematician is essentially free to take it or leave it and turn to something else, while an important problem in theoretical physics is usually a conflict, a contradiction, which 'must' be resolved. (…) The mathematician has a wide variety of fields to which he may turn, and he enjoys a very considerable freedom in what he does with them. *To come to the decisive point: I think that it is correct to say that his criteria of selection, and also those of success, are mainly aesthetical*. (1961, p. 2062; emphasis added)

More pronouncements like these can be found, but I will now move on to the second premise. It states that the modern mathematical concepts, originating in the subjective domain of our aesthetic sense, should not be effective in the objective domain of physics—and hence it is 'unreasonable' if they are. What does Wigner

[7]The selection of quotes is from Maddy (2007, p. 330).

[8]But, should one take into consideration their views on the matter, when they bothered to express them? My answer (for which I don't have space to argue here) is 'yes', but not everybody agrees; see Azzouni (2000, p. 224).

[9]Hamming continues by saying that "we have tried to make mathematics a consistent, beautiful thing, and by doing so we have had an amazing number of successful applications to the physical world" (1980, p. 83)—yet another expression of the Wigner problem.

claim here? The short answer is that what he says amounts, in essence, to voicing the generally accepted idea that there is no obligation for the world to conform to our human, parochial aesthetic preferences, in the sense that there is no obligation for the laws governing the world to be expressible in mathematical concepts. He expresses the same sentiment as Freeman Dyson who once asked, "Why should nature care about our feelings of beauty?" (1986, p. 103)

A more complete answer has to bring up the most intriguing element of the entire Wigner issue: yes, the Universe and the laws of nature are under no such 'obligation'—unless they, together with the human race, have somehow been *designed* to match. That is, unless a certain form of *anthropocentrism* is true. This means that we inhabit a 'user-friendly Universe',[10] that the human species has a privileged place in the grand scheme of things, that our subjective aesthetic inclinations (expressed in favoring certain concepts) have a correlate in objective physical reality,[11] i.e., are truth-conducive. This (intelligent) *design* suggestion has of course been long questioned, opposed, and considered 'unreasonable' by many. Thus, in doubting it in the second premise, Wigner doesn't in fact make any novel or controversial claim, but simply joins this rather influential line of thought. In the end then, although the premise may initially sound problematic, it turns out that it reflects the general naturalistic, agnostic (even atheistic) contemporary scientific zeitgeist.

We have now reached the third and last premise, which completes the argument. Under the (generally shared) assumption that modern mathematical concepts stem from the mathematicians' aesthetic inclinations, Wigner also pointed out that such concepts should not be effective in physics—and thus it is 'unreasonable' if they are. The third premise closes the circle by stating that they are so indeed: several physical theories proposed at certain points in time turned out to benefit significantly from the already developed mathematical theories and concepts. Let's ask, once again, what is claimed here. Similar to the second component (ii) of the first premise, one can't help but notice that this third premise also sounds like a factual claim; hence, such 'situations' can be documented historically. Then, the relevant question to ask here is whether Wigner (or anyone else) has done any *quantitative* assessment of the historical record, counting (α) the number of physical theories in this situation (the 'successes'), as well as (β) the ratio of successes to 'failures'.

I will now go over all available solutions, beginning with the one that takes issue precisely with the third premise. As it turns out, out of the six solutions I'll be discussing, only one attacks the first premise—although, as I'll argue in the last section of the paper, this premise is in fact the most vulnerable one.

[10]This reconstruction is heavily influenced by Steiner (1998), and I warn the reader that I may have been reading too much into Wigner's (1960) paper.

[11]Wigner makes the point about the independence (hence objectivity) of the laws of a huge variety of particular circumstances on pp. 4–5 in his (1960).

4 Revisiting the Available Solutions

The literature on Wigner's problem spans more than half a century, but (I dare saying) any attempt to deal with his argument is summarized by one (or several) of the six characterizations below (I present them here in abridged form, but more details follow):

It is *not* unreasonable to find mathematical concepts available to be applied in physical theories because …

> the situations when this happens are *not* numerous, and thus these 'successes' can be attributed to chance (Solution S1)
>
> Wigner's starting point, that mathematics is invented, is just *false* (S2)
>
> mathematical concepts have *empirical origins* (as opposed to aesthetical ones) (S3)
>
> applicability presupposes modeling, i.e., the 'preparation' of physical systems *in order to* apply mathematics (S4)
>
> there is *over-determination* in the relation between mathematics and physics, i.e., there is a lot of mathematics to choose from when we need to embed a physical insight (S5)
>
> our aesthetical mathematical sense is shaped by evolution, hence it is *sensitive to environment* (S6)

The first solution S1 attacks premise (3)—that several physical theories proposed at a later time turned out to benefit significantly from the application of mathematical concepts developed at an earlier time—by pointing out that its advocates make an exaggerated claim. Quantitatively speaking, the situation can be the result of pure chance.

Fair enough, the premise *is* rejectable in this way. However, before a counting of successes is done, the premise does not strike one as *clearly* false. One has to admit that the counting (once we settle upon *what* and *how* to count) *may* confirm Wigner. But this is perhaps too defensive; a sympathizer of Wigner's argument may also counterattack by proposing that this third premise should be read along *qualitative* lines too. Or, more precisely, that one should balance the strict quantitative reading against a more qualitative one. She could say that although there may be many natural phenomena that fail to receive a mathematical description, we should also judge the relative *relevance* of these 'failures' (and 'successes') in the larger context. From this perspective then, what the third premise says is that we should focus on the rather few *major*, truly *important* episodes in modern theoretical physics; and, if we do this, we'll see that they support Wigner's claim in premise (3). These major episodes are not very numerous (in absolute number) to begin with, hence the number of mathematical concepts and theories which were 'waiting', available for physicists to use them, should not be expected to be numerous either.

On this reading, the premise says that one can list *relatively* many major achievements in modern physics which fit Wigner's W_A scheme perfectly. It is, for instance, widely accepted that Einstein's General Theory of Relativity (developed, roughly, between 1905 and 1916) drew massively on Riemannian geometry

(developed before 1900), that quantum mechanics (in essence a product of the first quarter of the 20th century) makes essential mathematical use of complex numbers (among other concepts well-established before 1900). Moreover, groundbreaking work (within quantum field theory) on the classification of elementary particles by Wigner himself, Gell-Mann and others between (roughly) 1930s and 1970s employed group theory concepts (such as group 'representation') introduced much earlier (beginning with Frobenius, Lie, Shur, E. Cartan, and others.)

The list can be continued with several other well-known examples; thus, if the qualitative reading of this third premise is allowed to counteract the blind quantitative aspect, then one reads premise (3) as follows: 'restricting judgment to the *few* major breakthroughs in modern physics, many *of them* were anticipated by mathematical concepts and structures'. Read this way, one may begin to see that the support for this premise is actually not weak at all, to say the least.

The second solution S2 above rejects the first component (i) of the first premise. Such an objector denies that mathematics is 'invented' and, in particular, that modern mathematical concepts were invented to satisfy the mathematicians' aesthetic preferences. Thus, the picture Wigner proposes—that mathematicians invent concepts and decide to study those that foster beautiful theorems—is wrong. The correct metaphysical picture is something one may call Theistic Keplerian Platonism: there is a Creator of the Universe (God), and He made the world using a mathematical blueprint. For instance, when God created the solar system, He implemented in it the mathematical properties of the five 'perfect' solids—as the numerical values of the radii of the planets moving around the Sun; this is what Kepler, and others, genuinely believed.[12]

On this picture, a mathematician doesn't choose what concepts to study, but rather 'sees', with his 'mind's eye', what is 'there' (in the realm of mathematical forms) to be investigated; these concepts more or less 'force upon' him (to allude to the recent Platonist, Kurt Goedel). His job as a mathematician is thus not to invent anything, but rather to describe (as theorems), the eternal, true relations among these concepts; a mathematician's responsibility is therefore to discover the ways to connect these concepts, i.e., to discover proofs of the theorems. Moreover, once one does this properly, one may expect that one will *also* discover truths about the physical (material) world! This is so since, by assumption, these concepts and relations served as God's blueprints in designing the world.

Although some of the details of the story are still to be filled in, it should now be clear that there can be no problem as to why the Wigner-type coincidences arise. In fact, S2 is so radical that it almost generates an anti-puzzle: to the extent that one is a genuine mathematician (i.e., able to peek at God's blueprints), one *must* find such coincidences! Moreover, one should stop calling them 'coincidences', since they are the result of intentional acts of Divinity.

[12]A well-known passage from Kepler reads: "Thus God himself was too kind to remain idle, and began to play the game of signatures, signing his likeness into the world; therefore I chance to think that all nature and the graceful sky are symbolized in the art of geometry." Quoted in Dyson (1969, p. 9).

What about Theistic Platonism then? While it may have sounded credible a few centuries ago,[13] there are very few people who believe it today.[14] The general metaphysical view underlying it strikes us as creating more problems than it solves. (To mention an obvious one: what kind of evidence does one have, or *can* one have, for the postulations of this doctrine? What kind of epistemology do we have to develop to make sense of this picture of the world and of mathematics?) In the end, this seems a rather clear case in which the cure is worse than the disease, so to speak, as the amount of controversial metaphysical baggage one has to assume in order to make S2 work is *too* large. On balance, this way out seems then implausible; hence, one would be better off ignoring it, and looking for alternative ideas.

And it is not unusual that many find a more plausible alternative in the third solution S3. Thus, one can also object to the first premise, but not so much to component (i), as to component (ii). One rejects the view that mathematical concepts have aesthetical origins because, on this view, mathematical concepts have *empirical* origins. Philosopher Ernst Nagel's pronouncement is usually invoked here, as it nicely summarizes this position:

> It is no mystery, therefore, that pure mathematics can so often be applied (…) because the symbolic structures it studies are all suggested by the natural structures discovered in the flux of things. (1979, p. 194)

To this, a defender of Wigner's argument has two replies. First of all, this objector forgets about an important aspect of premise (1), namely that it is about *modern* mathematics, not about the basic, traditional arithmetical and geometrical concepts. They of course may well have empirical origins, and Wigner himself grants this in his paper.[15] But his point is not about these types of concepts; it's about the modern/advanced ones. These, as we saw above, are usually recognized as belonging to the corpus of mathematics in so far as the mathematicians find them interesting and intriguing, and not because they are 'suggested' by nature.

The second reply takes issue with the ambiguity of the idea that a certain structure is 'suggested' by nature. What this meant was clarified above in the form of conditions (a) and (b). Now, as is evident, many modern mathematical concepts and structures are the results of various kinds of generalizations and modifications of more basic concepts and structures, and this is just the normal course of mathematical development. If one grants that these basic concepts are directly reflected in nature (and thus one agrees that they were 'suggested' to the mathematicians in

[13]As Kline (172, p. 1028) describes: "the Greeks, Descartes, Newton, Euler, and many others believed mathematics to be the accurate description of real phenomena (…) [T]hey regarded their work as the uncovering of the mathematical design of the universe."

[14]A recent author seemingly embracing this idea is Plantinga (2011, pp. 284–91).

[15]See Wigner (1960, p. 2): "Furthermore, whereas it is unquestionably true that the concepts of elementary mathematics and particularly elementary geometry were formulated to describe entities which are directly suggested by the actual world, the same does not seem to be true of the more advanced concepts, in particular the concepts which play such an important role in physics."

this way), does it follow that the later modified/generalized concepts are *also* suggested by nature, via some kind of transitivity? Does it follow that if concept C* is a generalization/modification of concept C, and C is suggested by nature, then C* is also suggested by nature?

This reasoning is dubious; it is clear that one can modify and generalize a basic concept or structure in a multitude of ways, and yet, in perfect accordance with Wigner's position, the only generalizations/modifications that survive as mathematically viable are the ones which are regarded as 'interesting' enough to fascinate the mathematicians to further study them. So, although one can perhaps trace *all* modern mathematical structures to some 'natural structure', it is simply incorrect to maintain that this kind of transitivity supports the idea that modern mathematical concepts and structures are also suggested by nature.[16]

This fallacy is worth discussing in some detail, as I find it committed by the historian Kline and, following him, by the philosopher Maddy; see below[17]. In commenting on the modern developments of Group Theory, Kline points out (correctly) that the origin of this theory is in the attempts to solve polynomial equations, which he takes to be (correctly, again) "so basic a problem" (1980, p. 294), in the sense that solving equations which directly represent physical situations is an activity directly linked to ('suggested' by) the physical world. But one can accept this idea, and yet object to Kline's point (1980, pp. 293–4) that the more advanced concepts introduced in this theory much later (continuous symmetries, Lie algebras, group representations, etc.) *also* share this feature. In fact, they have nothing to do with 'nature' anymore. Unlike their ancestors, these concepts have been introduced for their aesthetical properties, as part of a mature and sophisticated mathematical theory.

Before making this point about group theory, Kline claims the same about Riemann's work on geometry. He says the following:

> The pure mathematicians often cite the work of Riemann, who generalized on the non-Euclidean geometry known in his time and introduced a large variety of non-Euclidean geometries, now known as Riemannian geometries. Here, too, the pure mathematicians contend that Riemann created his geometries merely to see what could be done. Their account is false. The efforts of mathematicians to put beyond question the physical soundness of Euclidean geometry culminated, as we have just noted, in the creation of non-Euclidean geometry which proved as useful in representing the properties of physical space as Euclidean geometry was. This unexpected fact gave raise to the question, since these two geometries differ, what are we really sure is true about physical space? This question was Riemann's explicit point of departure and in answering it in his paper of 1854 (Chapter IV) he created more general geometries. In view of our limited physical knowledge, these could be as useful in representing physical space as Euclidean geometry. In fact, Riemann foresaw that space and matter would have to be considered together. Is it to be wondered then that Einstein found Riemannian geometry useful? Riemann's foresight

[16] I discuss a different kind of transitivity in my (2012, Chap. 7).

[17] Ivor Grattan-Guiness reasons along the same fallacious line: "Much mathematics, *at all levels*, was brought into being by worldly demands, so that its frequent effectiveness there is not so surprising." (2008, p. 8; my emphasis).

> concerning the relevance of this geometry does not detract from the ingenious use which Einstein made of it; its suitability was the consequence of work on the most fundamental physical problem which mathematicians have ever tackled, the nature of physical space. (1980, p. 293)

But in a different work Kline himself contradicts this view:

> Bolyai, Lobatchevsky, and Riemann. It is true that in undertaking their research these audacious intellects had in mind only the logical problem of investigating the consequences of a new parallel axiom. (1964, p. 429)

So, it is after all unclear what Riemann "had in mind" when working on his new geometries: "only" the attempt to play with the mathematical possibilities,[18] or, as we were told above, his intention was in essence to solve "a fundamental physical problem", to find out "the nature of physical space".

I find the 'Riemann-qua-physicist' picture much less convincing than the 'Riemann-qua-pure-mathematician' picture. On reflection, bringing the former in discussion is perhaps the result of confusing two aspects of his work. One aspect has to do with understanding what he actually *did*. The question to ask here is: did Riemann's work consist in taking an element of physical reality (or an aspect of a physical theory of his time) and trying to describe it mathematically? Were his innovations 'suggested by the natural structures discovered in the flux of things'? Recalling clauses (a) and (b) above, the answer has to be 'no': there were no such things (i.e., differentiable manifolds) to describe in the physics of his time, let alone identified in nature, hence he couldn't have received any 'suggestion' from these two sources. (Slightly more precisely, what he was doing was to work out a mathematically profound generalization of the very idea of space.[19] This led to the notion of a differentiable manifold, and further, as part of the package, to a generalized notion of distance, together with a 'Pythagorean' theorem for such manifolds.) Physical-perceptual, tri-dimensional space provided of course the initial inspiration, and the object of description, for traditional geometry; nevertheless, as we saw, it just doesn't follow that devising ways to generalize it are also inspired or suggested by 'nature'.

The second aspect relevant here is what Riemann perhaps *hoped*, or desired, to achieve in his work—and this is an entirely different matter from what he actually did. It is perhaps true that Riemann hoped, even expected, that maybe one of the alternatives he was thinking up will be proven, as a matter of empirical fact, to be a description of the real, physical space (which, as we know, did happen in Einstein's work on General Relativity in 1916).

---

[18] In the same passage quoted above from (1964, p. 429), Kline calls this kind of work "an ingenious bit of mathematical hocus-pocus".

[19] In fact, the second passage in his 1854 masterpiece 'On the Hypotheses Which Lie At The Bases Of Geometry' contains the generalization point: "It will follow from this that a multiply extended magnitude is capable of different measure-relations, and consequently that space is only a particular case of a triply extended magnitude." (Riemann 1854; reprinted in Hawking 2007, pp. 1031–2; translated by W.K. Clifford).

With this distinction in place, talking about Riemann's "motivation" [as Maddy does (2007, p. 337)][20] is prone to perpetuate the conflation of the two aspects mentioned above. On one hand, we can of course assume that Riemann's 'motivation'—understood as *hope*—was to contribute to the progress of science by making available models of possible physical spaces. On the other, if 'motivation' refers to what actually led him to the manifold concept, we can be sure that he was *not* following some suggestion from 'nature'. He certainly couldn't have been in the business of taking cues from an extra-mathematical source ('nature', or physics) and writing down a mathematical formalism encoding them. To stress, what led him to introduce a new battery of concepts was the attempt to generalize, unify, etc.—and these are exactly aesthetical elements (recall, in a broad understanding of 'aesthetics').

Before we move on to S4, it is important to address a type of reaction very similar in spirit to the S3. Many are tempted to embrace a line of thinking of the *common cause* type, where the common cause is not 'nature' per se (as above), but the structural similarities, or symmetries holding both in nature and in the mathematical domain. Simplifying, what one often hears is this: 'Scientists study symmetries (structures, patterns) occurring in nature, and similarly, mathematicians are often fascinated by symmetries (structures, patterns) at an abstract level; thus, given this common basis, there is no surprise that a (temporal) harmonious correlation exist.'

To reply, one must repeat the idea that aside from some very basic symmetries (patterns and structures) studied in mathematics because they indeed pop up everywhere in the physical domain, the kinds of symmetries and structures making up modern mathematics are *invented* by the mathematicians. They are not imposed on them by 'nature'; they are selected (among the many available structures they invent) to be studied and developed because they are found aesthetically pleasing. A paraphrase of Dirac's point above illustrates the gist of this reply: the problem doesn't go away just because there is this common ground (symmetries, structures) between mathematics and physics:

> One may describe this situation by saying that the mathematician plays a game in which he himself invents the symmetries / structures while the physicist plays a game in which the symmetries / structures are provided by Nature, but as time goes on it becomes increasingly evident that the symmetries / structures which the mathematician finds interesting are the same as those symmetries / structures which Nature has chosen. (1939, p. 124)

Let us investigate the fourth solution S4 now. With it, we enter the region of less discussed objections, in part because these have been articulated rather recently. The essence of this solution is captured by the memorable words of Wilczek (2006a, p. 8): "One way to succeed at archery is to draw a bull's-eye around the place your arrow lands." The premise of W_A under attack here is the second one.

[20]Maddy (2007, p. 337) accepts the (problematic) Kline picture, backing it up with a quote from Kline himself (1968, p. 234): "So Riemann's motivation was not 'purely aesthetic' or in any sense 'purely mathematical'; he was concerned, rather, with the needs of physical geometry and his efforts were successful."

The critic points out that it is not unreasonable that modern mathematical concepts find uses in physics, since the applicability of mathematics relies heavily on *modeling*. That is, scientists don't apply mathematical concepts directly to nature, but they 'prepare' the physical systems first—they idealize, abstract, simplify them, etc.—precisely *in order to apply* the concepts they have available: mathematics typically applies to a *model* of the system. Thus, to adapt a well-known proverb: 'if all you have is a hammer, then turn everything into a nail'.[21]

This idea has been advanced by several authors, both philosophers and scientists. Among philosophers, Maddy (2007) and French (2000) argue for a similar view, admittedly troublesome for the Wignerian perspective. And yet one may insist that not all worries have been removed. For, on this picture, what we do is take a collection of aesthetically-generated concepts and model the physical reality to fit them. In doing so, we are successful quite often and—here is the crucial point in the rebuttal—one wonders, how could this happen? In this context it is instructive to reflect on what Jakob T. Schwartz (of the 'Ultracomputer' fame) writes:

> Mathematics (…) is able to deal successfully only with the simplest of situations, more precisely, with a complex situation only to the extent that *rare good fortune* makes this complex situation hinge upon a few dominant simple factors. (1986, pp. 21–2; emphasis added)

Is this 'good fortune' rare? The Wignerian demurs, and replies that in fact *many*[22] natural processes and phenomena "hinge upon a few dominant simple factors ", and thus remain meaningful after all the simplifications, idealizations, omissions are operated on them. So, why isn't it the case, one insists, that the opposite happens on a regular basis, namely that once we model and make these adjustments in order to apply mathematics, what we get in the end is so rigid and empty that a mathematical description, even correct, would be useless or meaningless? Looked at it from this angle, the Universe does seem 'user friendly' after all. So, to return to the proverb above, at the important junctures in science the scientists *often* tried to 'make' nails—and they succeeded. And, if they did, this means (the Wignerian insists) that this very possibility was somehow present there, and has to be accounted for.

In closing the discussion of S4, I should mention a variation on this theme by Wilczek (2006, p. 8), who writes:

> Part of the explanation for the success of mathematics in natural science is that we select what we regard as the scientifically interesting aspects of nature largely for their ability to allow mathematical treatment.

[21]Note that such an objector has no troubles to accept the first premise of W_A, that these concepts are in the mathematical corpus because they are interesting and intriguing; if this objection from preparation and modeling is viable, the origin of concepts is just irrelevant.

[22]Recall that 'many' is relative; it means, 'many *among the truly important ones*', since these are the ones that matter, as we remember from the discussion of premise 3 above.

That such a selection strategy[23] is in use in science sounds like a factual claim, and thus it is open to investigation (we can perhaps run a survey among scientists?). However, even before the results are in, the proposal strikes me as an exaggeration. To one who says that "[scientists] select what [they] regard as the scientifically interesting aspects of nature largely for their ability to allow mathematical treatment", a Wignerian is tempted to reply that although there may be cases like these on record (and Wilczek mentions one, the behavior of ultrapure semiconductor hetero-junctions subject to ultra-strong magnetic fields at ultralow temperatures), it is extremely hard to believe that this is a fundamental, and widely accepted, rule of the game in science. Moreover, when one recalls the standard examples of 'aspects of nature' to motivate the puzzle—as mentioned above: General Relativity, Quantum Mechanics, the characterization and classification of elementary particles in Quantum Field Theory—one remarks that *none* of these fit the selection strategy idea. Gravitation or the observed invariances holding among elementary particles were surely *not* "regard[ed] as scientifically interesting aspects of nature largely for their ability to allow mathematical treatment." On the contrary, it seems pretty clear that they were considered scientifically interesting *independently* of the existence of such a treatment.

The fifth solution S5 is built around the idea of what can be called *over-determination*. Now one rejects premise 2: there is nothing 'unreasonable' about the fact that the aesthetically-generated concepts and structures find a home in physics simply because there is a lot of them. There is a lot of mathematics to choose from when one looks for embedding a physical idea, and this quantitative fact alone solves the problem generated by the existence of the anticipatory coincidences.[24] To adapt Wilczek's archery metaphor, there is no mystery in the fact that when *many* arrows are shot, they'll eventually hit even a very small target.

The over-determination idea is admittedly very powerful, and yet can be doubted. To exploit the archery metaphor further, it is true that when many arrows are shot, they'll eventually hit even a very small target—but only if they are shot in the right direction, that is, in the direction of the target! How would that translate in less metaphorical terms: if the concepts are aesthetically-driven indeed, the fact that there are so many available to choose from doesn't really affect Wigner's overall point. That the 'arrows' (the subjective, aesthetically-selected concepts) are all shot in a direction other than the 'target' (the objective, careless world) surely doesn't make it more likely that the 'target' will eventually get hit *even if* there are very many of them.

[23]See also Maddy (2007, p. 339): "As a mathematical analog, I suggest that we tend to notice those phenomena we have the tools to describe. There's a saying: when all you've got is a hammer, everything looks like a nail. I propose a variant: if all you've got is a hammer, you tend to notice the nails."

[24]Maddy (2007, p. 341) puts it as follows: "With the vast warehouses of mathematics so generously stocked, it's perhaps less surprising that a bit of ready-made theory can sometimes be pulled down from the shelf and effectively applied."

Finally, we should now examine closely the sixth solution S6. Its key-insight is somewhat similar to S3, but its proponents develop the argument in a different manner. Supposing that one accepts that the modern mathematical concepts are generated and selected on the basis of our aesthetic sense (i.e., premise 1), Wigner's conclusion still doesn't follow since our aesthetic sense itself is a result of evolution, and thus shaped by, and sensitive to, our environment—i.e., to the 'natural structures' around us. Hence it enjoys some sort of objectivity, due to its origin; therefore the contrast subjective v. objective underlying the puzzle doesn't hold.[25] As I said, this solution re-iterates S3, but in a subtler way: it accepts the first premise at a superficial level, but rejects it at a deeper level.

Still, the Wignerian remains unconvinced. There is no doubt that evolutionary pressure plays an immensely important explanatory role in a variety of areas. Yet the explanation proposed here is hopelessly sketchy. On one hand, it's perhaps not hard to see how preferences for certain types of mating partners (the muscular and faster specimens, the more vividly colored ones, etc.) may be interpreted as reflecting what living creatures (humans included) take to be aesthetically pleasing. But, on the other hand, even if such a reduction of the aesthetical to the evolutionary advantageous is accepted (and it's by no means clear that it should be!), one still has a long way to go until one demonstrates that the aesthetical criteria *involved in shaping modern mathematics* are also subject to the same kind of reduction. When mathematicians talk in terms of beauty, they have in mind a highly formal, and abstract, type of beauty—not the 'corporeal', or mundane beauty (supposedly) efficacious in natural selection. The relevant question then becomes how exactly can the environment, and evolutionary pressure more generally, shape this formal/abstract beauty; this, the Wignerian urges, is a yet unanswered question. Moreover, given how counterintuitive the suggestion is after all, it is fair to say that the burden of proof (rather: answer) is on the proponent of this kind of solution.

5 A Sketch of yet Another Solution

At the end of the examination of the first solution S1 (the only one disputing premise 3), I concluded that there are ways to read this premise that would make it plausible. If all that matters is how we count the scientific episodes that vindicate Wigner, then we can also count in his favor; this is the 'qualitative' perspective I introduced above. The solution I'll be sketching now assumes this qualitative perspective, and still attacks this third premise, but in a more radical fashion. What I'll discuss here is not the number of successes of applicability [aspect (α) above], but the ratio of successes to failures [aspect (β)].

[25] I take Pincock (2012, pp. 184–5) to advance this line of thought: "Even an argument based on natural selection seems imaginable according to which our tendency to make aesthetic judgments is an adaptation precisely because these judgments track objective features of our environment."

The insight behind this solution is of an 'ecological' nature, i.e., it has to do with the way ideas 'survive' in the scientific 'environment'. The scenario I envisage is quite simple. Imagine a physicist—call him *Neinstein*—thinking up a novel physical theory at time t. Assume further that his idea is bold, and a candidate to belong among the several (yet, recall, not that many in absolute number) major scientific achievements one typically mentions in connection to the qualitative approach to the Wigner's problem. Now, it is a fact that such a scientist *has to* embed his insight into a mathematical formalism. I take it to be beyond doubt (again, as a matter of sociological/professional scientific fact) that without such an embedding, i.e., without the ability to write down the theory's central mathematical equations, the theory is extremely unlikely to draw any interest from the scientific community. Neinstein's theory—again, left in the form of a vision, or deep insight into the nature of things—might of course float around for a while, in the heads of other fellow scientists, but until it comes packed in a mathematical formalism, very few (if any) will be ready to take it seriously as more than mere speculation. In other words, the un-mathematized insight will not survive, just like organisms and species don't survive in uninhabitable environments. So, on this picture, what would have happened with a Neinstein proposing General Relativity before Riemann? Bluntly put, we would have never perhaps heard of him and his theory (or, as one might note, we shouldn't even call it a 'theory' in the first place, but stick with the initial label and call it a 'vision', 'revelation', 'insight', 'speculation', etc.)

In fact, one doesn't even need to make the assumption that our physicist's idea is a major one: *any* idea in physics, no matter how trivial, needs mathematical embedding. Yet, just to stay within the confines of the above qualitative interpretation of Wigner's point, let's assume we focus only on 'major' insights here. Thus, the thought behind this seventh solution is that *it is guaranteed* that for any case of a major idea in physics (developed at a certain time t), the scientist(s) proposing it *will have found* a mathematical formalism available to embed it, and thus express it as a proper scientific theory (where the formalism was of course developed, at least in part, at an earlier time t'). As intimated above, the reason for this is immediate: were this *not* the case, that idea would not have been recognized (as a valid scientific contribution); it would not have survived, and there would have been no theory to talk about today in the first place. Thus, there is no anticipation, no pre-adaption, no pre-established harmony, no miracle—only a filtering ('ecological') effect operating in the scientific environment. To stress the point above, if the mathematics had *not* been available when needed, such an idea/theory would most likely have been lost, perhaps forever—and thus the Wignerians could not have counted it (nor Neinstein) among the examples of (major) achievements/achievers.[26]

[26]Sometimes the physicists themselves try to develop the mathematics they need, but usually aren't successful. Here is the story of Gell-Mann in Steiner (1998, p. 93), relying on Doncel et al. (1987, p. 489): "[In trying to generalize the Yang-Mills equations] [w]hat Gell-Mann did without knowing was to characterize isospin rotations as a 'Lie Algebra', a concept reinvented for the occasion, but known to mathematicians since the nineteenth century. He then (by trial and error) began looking for Lie Algebras extending isospin—unaware that the problem had already been

To return to Weinberg's rendering of the Wigner problem, we now see that there *can't* be any cases of (major) achievements in physics in which the mathematician hasn't been 'there' before. The very fact that there is an achievement to talk about (i.e., recognized as such) is already a guarantee that there was a mathematician 'there' first. To begin by presenting a number of examples of achievements and then wonder how could it be that a mathematician was 'there' first is like wondering how could it be that all the people we find in a hospital are sick.

This new criticism[27] against the third premise amounts to maintaining that the quantitative comparison implicit in the third premise (even when read along the qualitative lines I proposed above) may actually be *unintelligible*. What we should be able to estimate is the number of the cases in which important physical ideas were advanced but *no* mathematical embedding for them was available—and then compare it to the number of successful cases (which, again, we assume we can list). Then, the argument goes, we have something to worry about only if the later number is much larger than the former (given premise 1, the aesthetical origins of mathematical concepts, and 2, the assumption of anti-anthropocentrism.) However, as I hope it is now clear, when it comes to this relevant ratio, we are able to (roughly) estimate only one number (the successes), but no way (even in principle) to estimate the other (the failures).

Acknowledgements Thanks are due to the stimulating audience of the *Models and Inferences in Science* conference, in particular to the editors of this volume, Emiliano Ippoliti, Fabio Sterpetti, and Thomas Nickles. The responsibility for the final form of the text is entirely mine.

References

Azzouni, J.: Applying mathematics: an attempt to design a philosophical problem. Monist **83**(2), 209–227 (2000)

Bangu, S.: Steiner on the applicability of mathematics and naturalism. Philosophia Math. (3) **14**(1), 26–43 (2006)

Bangu, S.: Wigner's puzzle for mathematical naturalism. Int. Stud. Philos. Sci. **23**(3), 245–263 (2009)

Bangu, S.: The Applicability of Mathematics in Science: Indispensability and Ontology. Palgrave Macmillan, London (2012)

Bourbaki, N.: The architecture of mathematics. Am. Math. Monthly **57**, 221–232 (1950)

Dirac, P.A.M.: The relation between mathematics and physics. In: Proceedings of the Royal Society (Edinburgh) (James Scott Prize Lecture), vol. 59, pp. 122–129 (1939)

(Footnote 26 continued)

solved by the mathematicians—but failed, not realizing that the first solution required eight components."

[27]A point distantly related to the present one is that there are major scientific achievements (the theory of evolution, and other work in biology) in which mathematics doesn't play any role (Wilczek 2007; Sarukkai 2005). However, Wigner's problem centers on physics (despite the general title of his paper).

Doncel, M., et al.: Symmetries in Physics (1600–1980). Servei de Publicacions, Universitat Autonoma de Barcelona, Barcelona (1987)
Dyson, F.J.: Mathematics in the physical sciences. In: The Mathematical Sciences (ed.) Committee on Support of Research in the Mathematical Sciences (COSRIMS) of the National Research Council, pp. 97–115. MIT Press, Cambridge (1969)
Dyson, F.: 'Paul A. M. Dirac' *American Philosophical Society* Year Book 1986 (1986)
French, S.: The reasonable effectiveness of mathematics: partial structures and the application of group theory to physics. Synthese **125**, 103–120 (2000)
Ginammi, M.: Structure and Applicability. An Analysis of the Problem of the Applicability of Mathematics. PhD Dissertation, Scuola Normale Superiore, Pisa (2014)
Grattan-Guiness, I.: Solving Wigner's mystery: the reasonable (though perhaps limited) effectiveness of' mathematics in the natural sciences. Math. Intelligencer **30**(3), 7–17 (2008)
Hamming, R.: The unreasonable effectiveness of mathematics. Am. Math. Monthly **87**(2), 81–90 (1980)
Hawking, S. (ed.): God Created the Integers. Running Press, Philadelphia & London, The Mathematical Breakthroughs that Changed History (2007)
Kline, M.: Mathematics in Western Culture. Oxford University Press, New York (1964)
Kline, M.: Mathematics in the Modern World. W. H. Freeman and Company, San Francisco (1968)
Kline, M.: Mathematical Thought from Ancient to Modern Times. Oxford University Press, New York (1972). (Vol. 3)
Kline, M.: Mathematics: The Loss of Certainty. Oxford University Press, New York (1980)
Maddy, P.: Second Philosophy. Oxford University Press, New York (2007)
Nagel, E.: 'Impossible Numbers' in Teleology Revisited. Columbia University Press, New York (1979)
Pincock, C.: Mathematics and Scientific Representation. Oxford University Press, New York (2012)
Plantinga, A.: Where the Conflict Really Lies: Science, Religion, and Naturalism. Oxford University Press, New York (2011)
Riemann, B.: On the Hypotheses Which Lie At The Bases Of Geometry. Reprinted in Hawking (2007) (1854)
Sarukkai, S.: Revisiting the 'unreasonable effectiveness' of mathematics. Curr. Sci. **88**(3), 415–423 (2005)
Schwartz, J.T.: The pernicious influence of mathematics on science. In: Kac, M., Rota, G.C., Schwartz, J.T. (eds.) Discrete Thoughts. Essays on Mathematics, Science and Philosophy. Birkhauser, Boston (1986)
Steiner, M.: The Applicability of Mathematics as a Philosophical Problem. Harvard University Press, Cambridge (1998)
Von Neumann, J.: The Mathematician. In: Newman, J.R. (ed.) The World of Mathematics, vol. 4, pp. 2053–2063. George Allen and Unwin, London (1961)
Weinberg, S.: Lecture on the applicability of mathematics. Not. Am. Math. Soc. **33**, 725–733 (1986)
Weinberg, S.: Dreams of a Final Theory. Vintage, London (1993)
Wigner, E.: The unreasonable effectiveness of mathematics in the natural sciences. Commun. Pure Appl. Math. **13**(1), 1–14 (1960)
Wilczek, F.: 'Reasonably effective: I. Deconstructing a Miracle', Physics Today, pp. 8–9 (2006)
Wilczek, F.: 'Reasonably effective: II. Devil's advocate', Physics Today, pp. 8–9 (2007)

Fast and Frugal Heuristics at Research Frontiers

Thomas Nickles

Abstract How should we model scientific decision-making at the frontiers of research? This chapter explores the applicability of Gerd Gigerenzer's "fast and frugal" heuristics to frontier contexts, i.e., to so-called context of discovery. Such heuristics require only one or a very few steps to a decision and only a little information. While the approach is somewhat promising, given the limited resources in frontier contexts, trying to extend it to fairly "wild" frontiers raises challenging questions. This chapter attempts to frame the issues (rather than to provide resolutions to them), and thereby to cast light on frontier contexts, which have been misunderstood by philosophers, the general public, and funding agencies alike.

Keywords Context of discovery · Decision under uncertainty · Fast and frugal heuristics · Frontier research · Gigerenzer

1 Introduction

Is it possible to make rational decisions in frontier research contexts, or must we be resigned to being non-rational or even irrational? Rarely can the more substantive decisions that must be made in such contexts satisfy the conditions of logic and probability theory required to meet the traditional standard of rationality. Since such decisions violate the standard, are they therefore irrational, no matter how careful are the researchers? Such an answer seems harsh, since investigators often can provide reasons for choosing B rather than A or C. In such cases these reasons are far from conclusive, yet they may be as well as one can be expected to do under the circumstances. To say that researchers blundered when a decision did not turn out well is simply unhistorical and whiggish if they were doing as well as can be expected in their frontier context (Kuhn 1970b, Sect. II).

T. Nickles (✉)
University of Nevada, Reno, NV, USA
e-mail: nickles@unr.edu

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_3

If the answer to the rationality question is a qualified affirmative, the next question is whether there can be an ameliorative role for an account of rationality in frontier contexts, that is, one that provides guidance to improve reasoning in such contexts, an account with normative force. Gerd Gigerenzer and the Adaptive Behavior and Cognition (ABC) Group give an affirmative answer to both questions (Gigerenzer et al. 1999; Todd and Gigerenzer 2000). They want to extend their account of "fast and frugal" heuristics and "ecological rationality" to real-world contexts characterized by uncertainty, and they want to improve human reasoning in both routine and risky contexts.[1] They reject the invidious use of context-of-discovery/context-of-justification distinctions by the logical empiricists and Karl Popper to exclude frontier research from epistemology and scientific methodology. They dismiss Popper's Romanticism (theories as non-rational creations of the imagination) and return to a more Enlightenment conception of reasoning as rule-based calculation or computation—but not to the Enlightenment, universal, a priori standard of rationality. In many cases, they claim, fast and frugal (f&f) heuristics will do as well or better than traditional rational decision theory, where the latter is applicable.

One of their goals is to extend their core results to new areas, both descriptively and normatively. Frontier research is one of these areas. Given the limited informational resources of frontier research, exploring the promise of the ABC approach would seem to be well motivated. A more positive motivation is provided by the many studies showing that experts typically use less information than novices and by the many stories of scientists and technologists who made breakthroughs by ignoring one or more theoretical constraints or even empirical data. In these cases we have the irony that the way ahead in dealing with the problem of sparse information may be to make the sparse information even sparser!

Failure at the frontier would not be a refutation of the program as a whole, which has already enjoyed many successes. Todd and Gigerenzer themselves remark that "Some higher-order processes, such as the creative processes involved in the development of scientific theories or the design of sophisticated artifacts, are most likely beyond the purview of fast and frugal heuristics" (2000, 740). But frontier research involves many things, and the question is worth investigating.

Accordingly, this chapter is a first attempt to explore whether or not the f&f, ecological rationality program can be usefully extended to fairly "wild" frontier research contexts. As Gigerenzer and his many associates have shown over the past thirty years, the approach works very well in several important practical contexts that are not frontier-like. When it comes to frontiers, Gigerenzer's main contribution is

[1]There are other lines of both descriptive and normative work of this sort. Meehl (1954) was critical of the expert-intuitions approach typical of traditional clinical psychology and argued, with empirical support, that relatively simple decision rules often provide better results. Meehl's rules are not always fast and frugal, and Gigerenzer does not reject expert intuition as a sometimes-reliable basis for decision-making. See also Dawes (1979). Bishop and Trout (2005) argue for an ameliorative epistemology to replace standard analytic epistemology. They identify with the Meehl tradition and are rather critical of Gigerenzer's approach (Chaps. 8 and 9).

the "tools-to-theories" heuristic, by means of which the tools that researchers use to study organisms are subsequently projected onto the agents themselves as the explanation for how their cognitive systems work. Given limitations of space, I save detailed discussion of this heuristic (and also case-based thinking) for other work (Nickles draft-2015).

2 The Problem of Research Frontiers

It is at the frontiers of inquiry in the sciences, technologies, and the arts where our most substantive knowledge about the universe, including us humans, grows. The central problem of frontier theory of inquiry is this. (1) Roughly speaking, the wilder the frontier, the sparser or more badly organized the available domain knowledge. (2) Yet for the heuristics and other procedures employed, there is a roughly inverse relationship between problem-solving power and specificity, or problem-solving power and domain knowledge, just as turned out to be the case in computer science.[2] (3) Reasoning in frontier contexts involves a good deal of decision-making under uncertainty, often-extreme uncertainty, not merely decision-making under risk. Thus standard probabilistic methods, including standard Bayesian analysis, will not suffice.[3] (4) While past frontiers have been tamed by successful science, their problems now being reduced to routine solution methods, current frontiers are just as challenging for us as past frontiers were to the scientists who experienced them. As Hume argued intuitively, and as Wolpert and Macready (1995), Wolpert (1996) have argued in a series of "No Free Lunch Theorems," no inductive rule can be known a priori to be superior to any other when all possible domain structures are taken into account. And frontier research is that which ventures into new domains.

As science expands into new territory, including more deeply into an established field, there is no guarantee that old techniques will continue to work. Moreover, ongoing developments frequently undermine deep orthodoxies of the past. Consider what happened to remarkably successful classical mechanics when scientists attempted to extend it to the deeper atomic-molecular domain. Thus cumulatively building upon past work can take scientists part way at best—and sometimes into a blind alley. Science is nonlinear in the sense that even a fairly normal result may produce a conflict with another claim than cannot be resolved. And even when scientific change is more evolutionary than revolutionary, the change can be as dramatic as you please, given enough time. To be sure, ongoing frontiers can be tamed a bit by increasing expertise and by technological advances (e.g., computer modeling versus modeling the old fashioned way), but only a bit. If they could be

[2]See the quotations from Edward Feigenbaum and Allen Newell in my (2015).

[3]There are exceptions, e.g., causal Bayesian networks and (other) algorithmic searches of large databases.

tamed to the level of completely routine science, they would not be wild frontiers or even *research* in the fullest sense, and the growth of that specialty area would be minimal.[4] As our techniques improve, our aspiration levels rise.

At stake is the big issue of whether a "frontier epistemology" or "frontier theory of inquiry" is even possible. There is, of course, no question of finding a turn-the-crank algorithm that routinely produces novel successes, let alone a universal logic of discovery.[5] Even requiring that frontier decisions reflect reliable processes (as required by reliabilist epistemology) may be too strong.

How *do* creative scientists make decisions concerning their next research steps at frontiers of research? Skilled people of all kinds are creatures of routine, but here I am talking about non-routine decisions of the sort that determine the direction of the research or whether or not to explore previous results or practices in a substantially transformed manner. I shall not here attempt a taxonomy of frontier contexts, although I think a large collection of concrete examples would be useful and might even lend itself to a limited degree of higher-order classification, at least at the level of problems. Significant research often requires developing new tools and/or new resources, material, conceptual, practice-based, and organizational. It is frequently a dynamic enterprise that involves developing new vocabulary, among other tools, and articulating new or shifted goals. A piece of pragmatic wisdom is that there are two ways to solve a problem: You can get what you want or you can want what you get. If you know exactly what you need to get, you are doing fairly routine, normal science. To the degree that you are open to wanting what you get, you are entering a frontier.

My general answer to the question of how scientists make decisions in risky and in uncertain research contexts is by means of heuristic appraisal, i.e., evaluation of the heuristic potential of the different research steps considered (Nickles 2006). My ultimate aim is to make this idea crisper by linking it with extant studies of heuristics, in this case Gigerenzer's. In this chapter I can make only a few gestures in this direction.

What are f&f heuristics? They are computationally precise rules that lead to a decision in one, two, or a very few, clearly defined steps and that do not require large amounts of information. A fast heuristic is one that does not require much computation. Typically, its search is limited to a small amount of the total information available.

But how useful Gigerenzer's f&f rules can be in frontier contexts remains a challenging question. Most of the successful applications to date lie in fairly routine sorts of decision contexts. Moreover, there is an ambiguity in talk of domain specificity (specificity of application versus specificity in explicitly building domain

[4]For a defense of these claims, see Nickles (2006, 2015, draft-2015).

[5]Strong proponents of Bayesian methods sometimes leave the impression that Bayesian methods are the updated form of a universal, content-neutral scientific method; but many frontier contexts would seem to pose severe difficulties for Bayesian methods as for other approaches.

information into a decision rule or an algorithm itself: Sect. 6 below). And some authors loosely lump together risk and uncertainty, often in popular writings. Like the ABC Group, I shall follow the standard usage of decision theorists and economists in distinguishing decision making under uncertainty from decision making under mere risk. While frontier contexts are certainly "risky" in a popular sense, it is uncertainty that creates the deep difficulties.

What exactly is the difference? According to standard decision theory as employed by economists and others, decision-making under risk is a matter of maximizing one's expected utilities. In the full-blown case, one has (a) several action options, (b) a partition of relevant future states of the world, (c) a probability distribution over them, plus (d) utility assignments that reflect a well-defined preference ranking. By contrast, some of this information is missing from situations of decision-making under uncertainty. In frontier research contexts, one will rarely have a partition of states of nature and their probabilities, and one is likely to be unclear about the ultimate goal and hence the preference ranking and utilities. Uncertainty in a broader sense also characterizes the consequences of present work in the context of other results, in terms of the degree to which they may be destabilizing or unifying. Sometimes even fairly normal results turn out to be destabilizing, e.g., by introducing discrepancies ("anomalies") that, despite much effort, can never be reconciled with extant theory or practice.

I emphasize that the frontier sparseness problem is not always a matter of a small quantity of domain information, e.g., empirical data. Even more important, in most cases, is the structure of the domain, which has to do with how domain information is organized in terms of lawful principles, causal relations, statistical correlations, and the like. Today everyone is talking about problems of handling Big Data, how to mine it for relevant information, how to discern patterns in it, etc. What I like to call *the knowledge pollution problem* can present major difficulties even in frontier contexts. Today we are so awash with information that it is often difficult to identify the relevant pieces. The answer that we seek may be hiding in plain sight. No physicist has the time to read *The Physical Review*, for example, now that it has gotten so large. Ditto for chemists attempting to read even chemical abstracts. Chemical Abstracts Service (CAS) has now registered almost 100 million distinct organic and inorganic substances, a good many of which are being studied in new research papers each year.[6] The periodical *Chemical Abstracts* grew so large that the American Chemical Society stopped print publication of it in 2010.[7] The society developed SciFinder as a tool to recall relevant items in the huge chemistry database. Research scientists have developed heuristic filters of various sorts as parts of their individual research styles, but most of them surely employ keywords as a literature search procedure that is much faster and more frugal than reading

[6]Moreover, as the CAS website informs us, CAS deals only with "disclosed" chemistry, not the undisclosed, secret research for military and proprietary purposes.

[7]*Wikipedia* article "Chemical Abstracts," accessed 2 June 2015.

abstracts. In fact, a f&f heuristic filter here would be: If an article displays one of your research keywords, read the abstract! A slightly more complex rule would prioritize according to the number of keywords from your list. Perhaps these are not successful f&f heuristics in Gigerenzer's sense, since the proportion of useful hits is not likely to be high, but we might call them f&f neighbors of those rules.

In *The Structure of Scientific Revolutions* (1962, 1970a, b) Thomas Kuhn drew a sharp distinction between normal science and extraordinary or revolutionary science, but it is more plausible to think of a spectrum of frontier possibilities between these extremes. At the "normal" extreme the frontier is pretty tame: the goals and problems are fairly well structured, to use Herbert Simon's term (1977, Sect. 5.3). At Kuhn's revolutionary extreme the frontiers become so wild that Kuhn spoke (problematically) of incommensurability between the new and the old practices. Be that as it may, in fairly wild frontiers the problems are ill structured, and even the goals and vocabulary are likely to shift as the research proceeds. I attempted to capture the flavor of moderately wild research frontiers in Nickles (2015). Here I shall add a quotation from Richard Rorty. Although he is not usually quoted *in extenso* by philosophers of science, the following extended passage characterizes the revolutionary extreme as I conceive it.

> The gradual trial-and-error creation of a new, . . . vocabulary—the sort of vocabulary developed by people like Galileo, Hegel, or the later Yeats—is not a discovery about how old vocabularies fit together. That is why it cannot be reached by an inferential process—by starting with premises formulated in the old vocabularies. Such creations are not the discoveries of a reality behind the appearances, of an undistorted view of the whole picture with which to replace myopic views of its parts. The proper analogy is with the invention of new tools to take the place of old tools. To come up with such a vocabulary is more like discarding the level and the chock because one has envisaged the pulley, or like discarding gesso and tempera because one has now figured out how to size canvas properly.
>
> This Wittgensteinian analogy between vocabularies and tools has one obvious drawback. The craftsman typically knows what job he needs to do before picking or inventing tools with which to do it. By contrast, someone like Galileo, Yeats, or Hegel (a "poet" in my wide sense of the term—the sense of 'one who makes things new') is typically unable to make clear exactly what it is that he wants to do before developing the language in which he succeeds in doing it. His new vocabulary makes possible, for the first time, a formulation of its own purpose. It is a tool for doing something which could not have been envisaged prior to the development of a particular set of descriptions, those which it itself helps to provide. (1989, 12–13)

So one way to put the larger question behind this chapter is whether there can be an ameliorative epistemology of the frontier. My question is whether and to what extent Gigerenzer's approach can be extended from contexts of decision-making under risk to decision-making under uncertainty. Mine is the meta-heuristic exercise of exploring the promise (positive and negative) for extending his "less is more," f&f heuristic approach explicitly to frontier research contexts. I shall suggest that introducing some broadly parallel Kuhnian elements may help this enterprise, by adding some desirable concreteness to the treatment of empirical scientific research.

3 Some Background: Kuhn, Simon, Gigerenzer

In my view the character of frontier research has been misconstrued by many philosophers of science as well as by the general public and funding agencies. This is surprising, since the frontier is the primary growth-point of knowledge. Karl Popper stated that the central problem of philosophy is the problem of the growth of knowledge (Popper 1972). But, as is well known, Popper himself, like the logical empiricists, insisted that there is little that is philosophically interesting to be said about this, that predictive testing of ideas already on the table is where the philosophical action is. Two reasons in support of Popper's position are that (1) there is no single "nature" to the frontier: we must reject any trace of frontier essentialism, and (2), accordingly, "the scientific method" that defines science is a myth (*pace* his emphasis on falsifiability as a criterion of demarcation).

I believe that Popper was right on both of these counts but that there is nonetheless important epistemological/methodological work to be done with frontier contexts. One attempt was Kuhn's. In *The Structure of Scientific Revolutions* (1962) Kuhn aspired to rehabilitate context of discovery as a legitimate domain for philosophical investigation. In that and related papers from this period (e.g., 1970a, b), he broadened the context beyond logical relations to include rhetorical relations such as similarity and analogy and argued that much problem solving in science is based on noting similarities of the current puzzle to a paradigm-defining collection of exemplary problems and solutions. Second, he stressed the importance of heuristic fertility over representational truth in decision-making at the frontier. Although there is much to criticize in Kuhn's dynamical model of scientific change, I believe that both these moves are on the right track in opening up frontier contexts to investigation.

The positivist-Popperian position was also challenged from a very different quarter. While many philosophers were still relying on the two-context distinction to dismiss the idea of frontier epistemology, by the late 1950s Simon and company were busy inventing artificial intelligence, much of which was devoted precisely to just that—what we may call frontier epistemology. For Simon and colleagues, attention to the discovery context was not only possible but also necessary if AI were to advance. Simon stressed the centrality to inquiry of both problem solving and heuristics. From around 1960 his insight was that, contrary to the naysayers, we already had a good deal to say about the seemingly esoteric context of discovery. For scientific discovery is a form of problem solving, we already knew a good deal about that, and ongoing research was already teaching us more. Specifically, problem solving (including problem finding) is search, and studying the advantages and disadvantages of various types of searches through different kinds of problem spaces is both a possible and a fruitful area of investigation. The proper use of

heuristics makes a tremendous difference in the efficiency of such searches.[8] In short, Simon and AI made the treatment of discovery both unavoidable and tractable.

Roughly speaking, Simon's important distinction between well-structured and ill-structured problems parallels Kuhn's distinction between research puzzles in normal science and the deeper, poorly defined problems central to revolutionary science. Well-structured problems are those with well-structured problem spaces, which means, basically, that there is a clear decision procedure for determining whether a given point in the space represents an adequate solution.

As for heuristics, Simon's were at first domain-neutral heuristics such as hill climbing, backward chaining, and means-ends analysis (Newell and Simon 1972). However, he soon recognized that these were helpful only in certain kinds of empirically contingent circumstances. In that sense they were not purely a priori, universally applicable rules but had implicit empirical content of a fairly general sort. After all, they were heuristics![9] But meanwhile, AI adepts were discovering that expert systems and knowledge-based computation in a stronger sense, by featuring heuristics that explicitly incorporated domain content, could address more difficult problems than the General Problem Solver of Newell and Simon, although they were thereby far more restricted in scope of successful application.[10]

Simon is surely most famous for challenging the traditional, a priori conception of rationality, replacing it with his positions on bounded rationality, satisficing, heuristics, and contextual (environmental) sensitivity. Thus there is a premium on economy of research. Attempting to do too much with the limited resources available is not cost-effective.

Here again there is something roughly parallel in Kuhn. Both men make an important contribution to a pragmatic account of how science works. Recognizing how far successful scientific research departed from the logical models of philosophers and from pious claims about the search for truth, Kuhn famously (or notoriously) reinterpreted scientific progress as a matter of moving fruitfully away from, or beyond, a starting point (a *terminus a quo*) rather than as moving toward a final goal of representational truth about that domain of the universe (a *terminus ad quem*). An important implication is that, in decision-making at frontiers, fertility trumps truth. We thus avoid the "to the max" philosophy of aiming for final, representational truth. Even when scientists think they know upon which path truth lies, they will typically choose a more fruitful path, when available, one that gives

[8]AI researcher Douglas Lenat soon extended these ideas to all of science. "Discovery is ubiquitous," he said (Lenat 1978). Problem solving as search pervades scientific work, including the testing and justification phases (regarded as ongoing practices or processes rather than finished products), and is not limited to an early stage of "context of discovery." Thus understanding discovery is necessary to understand science.

[9]Wimsatt (1980) stressed that even reliable heuristics work well only in limited domains of application.

[10]The distinction is between those AI systems that incorporate rules gleaned from human experts and knowledge-based systems more generally.

them something more doable or that promises new and interesting results. For instance, several physicists expressed disappointment in the apparent discovery of the Higgs boson at CERN, on the ground that a deviation from expectations would have left them with a lot of interesting physics yet to do. But in fact, when it comes to general theoretical claims (and some other claims as well), the absolute truth is not attainable with any assurance—and especially not in frontier research. There is no direct test for truth, either in advance or post hoc, that is, after the research is completed. By contrast, we can know, post hoc, whether or not a line of investigation has been fertile. The question is: how reliably can we estimate this in advance? How reliable can heuristic appraisal become?

As already hinted, this may be good news for the progress of inquiry but bad news for strong scientific realist positions, for there is no reason to think that truth (or truthlikeness) is correlated with fertility in the human order of inquiry. Fertility is not a reliable proxy for truth. Indeed, when working scientists speak of truth, many of these locutions are best interpreted as a summary statement concerning fertility. To use medieval terminology, what is fruitful in the *ordo cognoscendi* (order of knowing) may yet be very far from the *ordo essendi* (order of being or causal order). From my historicist point of view, the closer we look at the processes of scientific investigation, including the history of science, the less attractive the philosophical position of strong, convergent, philosophical realism looks in those areas where we lack a high degree of controlled experimental access. Conversely, strong realists tend to treat frontier research in a shallow manner, as tamer than it usually is (Nickles in press).

4 Gigerenzer's Ecological Rationality

Gigerenzer and his associates in the Adaptive Behavior and Cognition (ABC) Group locate themselves in the Simon tradition of ecological rationality. The f&f heuristics program is an important development in empirical decision theory, what we might call "behavioral decision theory") as a kind of generalization of behavioral economics.[11] "Loosely speaking," writes, Gigerenzer (2010, p. 50), "the heuristic lets the environment do much of the work." Context-sensitivity here means taking the structure of the environment into account, thereby avoiding what is often called "the fundamental attribution error," which, by failing to take into account environmental factors, forces us to attribute unnecessary cognitive complexity to the organism.

The f&f heuristics approach is a bold departure from traditional decision theory. The latter requires maximizing expected utilities and thus typically requires large

[11]Gigerenzer's treatment of heuristics thus differs from the "heuristics and biases" program of Tversky and Kahneman, which retains the classical conception of rationality (Kahneman et al. 1982). See, e.g., Gigerenzer et al. (1999, 2011).

amounts of information (often unavailable) and huge amounts of computing power, often impossible amounts. A f&f heuristic must be fast (one, two, or perhaps three steps), frugal (employ a minimum of informational content), and be computable. To be computable the heuristic must contain a precise search rule, a decision rule, and a stop rule. In sum, the goal to be achieved and the steps needed to achieve it are completely determinate.

Gigerenzer's breakthrough came when, surprisingly, the Germans he was testing performed better on a population comparison of pairs of U.S. cities than Americans themselves did. The Germans could be interpreted as using a simple recognition heuristic: If you recognize the name of only one of the two cities, choose that one as larger. The best performers were those agents in the Goldilocks position of knowing something about American cities (in this case) but not too little or too much. The Americans did poorly, on average, because they knew too much: they had heard of all of the cities.

Gigerenzer and the ABC Group have since extended this line of research to several f&f heuristics in a variety of applications, including medicine and law, showing that they often do as well or better than far more computationally expensive methods such as multiple regression. Gigerenzer does not dismiss traditional statistical methods completely, however, granting that there are contexts in which they give better results. One of Gigerenzer's main points is that there is no "one size fits all" method. Accordingly, he and the ABC Group speak of an "adaptive toolbox." Which tools are useful depends upon the domain of application, the ecological context.

5 Some Reasons for Optimism

In the remainder of this contribution I first itemize some reasons for thinking that something like the ABC approach might help us to understand decision-making in frontier contexts, and then I list some challenges that such an extension faces.

(a) Many empirical studies show that experts search for and use less information than do novices. Experts are far more selective. True, at far frontiers, no one is an expert in terms of detailed domain knowledge, but there remains a big difference in research know-how.
(b) Using frugal heuristics of some kind is the default situation for many frontier decisions, given the scarcity of domain information.
(c) Being quick, easy, and cheap to use, f&f heuristics are potentially valuable, especially in the resource-poor, early-to-middle stages of a research project, when researchers know something, but neither too little nor too much.
(d) Not even wild scientific frontiers are entirely new. To some degree old methods of investigation can be recycled or adapted to the new application. Typically, many tools are in play, some old, some newly evolving. For instance, a new instrument may produce reams of new data that can be

analyzed by older methods, and a new model will often permit familiar kinds of computation.

(e) The method of hypothesis, long taken to be the central method of science, already retains the less-can-be-more motif, since the researchers introduce hypothetical premises into key scientific arguments, e.g., for purposes of potential explanation and prediction. Contrary to Euclidean methods that require reasoning only from already established truths, or to Baconian methods that require those truths to be observational from rich data sets, the method of hypothesis both dispenses with the known-truth requirement and flourishes in data-sparse domains. It even functions to identify which observations are most relevant to seek. (However, in a sense it adds new information, hypothetically.) Modeling methods push still further by allowing premises already known to be false. Modeling is the use of heuristics writ large.

(f) Heuristic reasoning typically works by neglecting some factors. For example, a back-of-the-envelope calculation may reveal the heuristic promise of a line of investigation where rigorous accounting for the variables in play or where immediately requiring precise agreement with exact measurements would spoil the exercise (see, e.g., Miller 2002).

(g) There is surely at least a grain of truth in Levins (1966) famous argument that, when dealing with complex systems, scientists must strike a compromise. Levins prefers to sacrifice quantitative precision to generality and realism. And then there are the many instances in which experts from other fields come in and make a breakthrough partly by ignoring the constraining expert culture or "context bias" of the field in question, while perhaps introducing new modeling ideas or constraints of their own.

(h) Much frontier thinking is rhetorical, based on analogies, similes, or metaphors. All of these make some features salient while ignoring others. In their introduction to *The Psychology of Problem Solving*, Pretz et al. (2003, 10–11) remind us that, when faced with analogy problems, "knowledge is actually an impediment to problem-solving success." Good modeling typically depends on identifying apt analogies while not becoming distracted by disanalogies. "The man who knows too much" can easily think of disanalogies that, for him, constitute knock-off arguments against a promising idea or technique. A key to good modeling, especially in Kuhnian matching of current problems to exemplars, is to be able to resolve the current problem into a combination of exemplary problems, much as a legal expert can resolve a complex case into multiple dimensions, each of which has legal precedents. Relevance does not require identity.

(i) In their article "How forgetting aids heuristic inference" Schooler and Hertwig (2005) argue that forgetting can be functional for higher cognition, as in the recognition heuristic, when knowing too much spoils performance. There is much anecdotal information that taking a break from work or changing the venue or even closing one's eyes (Vredeveldt et al. 2011) can help us gain a perspective on essentials, by suppressing the clutter of detail. We have all had

the experience of giving some of our most insightful lectures when arriving at class underprepared and "winging it."

(j) What of "follow your gut instinct" and "trust your intuition" as f&f heuristics for experts working at frontiers? These fit the less-can-be-more idea in the sense that they are immediate, one-step procedures that require little or no conscious deliberation. I agree with much that Gigerenzer writes in his *Gut Feelings* (2007) and also with Damasio (1994), Thagard (2004, 2008) and others that emotion can be cognitively important and can play a guiding role in good decision-making. We believe that much animal cognition employs the emotions as an important adaptive endowment. We also want to recognize the role of skill or expertise. But caution is necessary. Apart from the difficulty of determining whether or which subconscious rules are in play (see (ix) in Sect. 6), there is the Meehl tradition that argues that simple statistical rules often outperform the intuitions of supposed experts. (Is there thus something of a tension within Gigerenzer's own view?) Furthermore, intuitions that are the product of deeply ingrained habits of experts are again the products of much domain experience (either ontogenetic or phylogenetic via evolutionary learning) and thus reliably govern behavior only in normal situations rather than in unfamiliar, frontier contexts.

(k) Given that f&f heuristics tend to work well within a specific environmental niche or *Umwelt* with its domain structure, and given that cognitive economy is an adaptive advantage in the biological world, we should expect evolution to have hit upon many f&f heuristics that reduce cognitive load well below what traditional logic and statistical inference requires, e.g., in adapting the critter to the affordances crucial to the organism's lifeways (Gibson 1979; Norman 1993). On the ground of evolutionary continuity, we should expect some of these to be wired into us humans as well, as the ABC program supposes. However, again, the products of evolutionary adaptation will surely be regularized behavior patterns rather highly specific tricks for faring well in unfamiliar domains. Besides, regularized action rules would make the critter vulnerable to predators, who will learn to recognize the behavior patterns—unless some sort of randomizing decision procedure is built in.

(l) Quite generally human inquiry proceeds in a broadly evolutionary manner, by default, for a variation-selection process is the only one we know that solves the Meno problem (Nickles 2003). It is perhaps debatable whether evolutionary *biological* processes themselves employ heuristics, since it makes little sense to attribute prospective, problem-solving purposes, etc., to them (unless we count such things as exaptation as heuristics: see Wimsatt's commentary to Todd and Gigerenzer 2000); but as Simon and others have emphasized, heuristics become crucial in the human case. Lacking the bountiful resources of Mother Nature (every organism on earth as a field experiment in a vastly parallel "computation"), we need to take advantage of our small measure of lookahead to employ a more efficient variation-selection process, one that we can point in promising directions.

Alan Turing once observed that an intelligent computer need not know what arithmetic is, because its behavior is algorithmic. Similarly, Mother Nature does not know what arithmetic or prey or predator or mate is but has nevertheless succeeded in creating "endless forms most beautiful" (in the words of the final paragraph of Darwin's *Origin*). If frontier research were governed completely by fast and frugal heuristics, operating at the Turing or Darwin limit, it might be able to make progress, but it would be extremely slow progress. As noted, we human beings do have a bit of lookahead and can formulate and address problems off-line, so to speak. The history of the sciences and the arts discloses that we can proceed at a relatively fast pace, so some degree of improved efficiency is possible. If this way of thinking is correct, then Gigerenzer's f&f heuristics may have some purchase in human frontier research, but other, more powerful heuristics based on "knowing arithmetic" (i.e., domain expertise) may be more important.[12]

Less-can-be-more heuristics are the key to solving the Meno paradox of inquiry, as Simon and others have noted. Paradoxically stated, a less-can-be-more approach is necessary to get more from less! The Meno paradox concludes that learning—getting more knowledge from less, more design from less—is impossible, because the inquirers either know already that which they seek (and hence genuine inquiry is impossible) or else would not be able to recognize the solution even should they stumble upon it accidentally (whence, again, successful inquiry is impossible).[13] The key to solving the paradox is to avoid this all-or-nothing dilemma by going between the horns, by using indicators that tell us, perhaps highly fallibly, whether we are getting warmer or colder in the search for a solution. And, regarding research as a slow, step-by-step process of variation and selection, the satisficer will be content with a solution that meets the aspiration level "heuristically suggestive" rather than "certifiably true."

It was Darwin's theory of evolution that opened the door to the large domain of variation-selection processes that show how to get more from less, in Darwin's case more biological design from less. My own view in Nickles (2003) is heavily indebted to Campbell (1974) and Dennett (1995) as well as to Simon (1945, etc.) and Wimsatt (2007). If universal evolutionists Campbell and Dennett are right that all novel design emerges from a BVSR process, the questions become: (1) Which ones are available in which kinds of frontier situations? and (2) Which ones among the first set are more efficient than others for a given task?

(m) In order to have a guiding role at the frontier, an ameliorative role, a heuristic must satisfy some conditions, but the requirements should not be set so high that they are impossible to meet. Since it is a form of appraisal, heuristic

[12]On Darwin and Turing see Dennett (2009). For his Tower of Generate and Test see his (1995, 1996).

[13]According to Simon (1992, 155): "Intuition is nothing more and nothing less than recognition." Klein (1999) develops a "Recognition-Primed Decision" model in which intuition plays an important role.

appraisal is normative; but it need be only weakly so to have some purchase. There can be no requirement that the heuristic provide a route to truth or that it meet the conditions of traditional rationality theory or even reliability theory. As satisficers we must be content with highly fallible indications of fertility. Often we must make do with something weaker than f&f rules. In the weakest case we might fall back on pragmatic vindication: a heuristic has occasionally seemed to work, we must act now, and nothing better is available.

Once we adopt Gigerenzer's wider concept of rational behavior (jettisoning the rational decision theory textbook statements about irrational behavior), we have a better chance of making sense of frontier inquiry.[14] But extending the f&f approach to frontier contexts will require still more flexibility, and that move will generate new challenges of its own (Sect. 6). Even if helpful f&f rules in research contexts do not fully meet Gigerenzer's requirements, this would still be progress, with Gigerenzer's model being the limiting case. Stated otherwise, our satisficing in frontier contexts may require treating the Gigerenzer limit as a rarely realizable optimum.

(n) Satisficing applies to content-specific research *goals* as well, and hence to problem formulations. Creative researchers in fairly wild frontier contexts remain open to unexpected twists that may alter their research goals. So a f&f heuristic in the form of a production rule might be: If a research procedure produces a surprising result, then decide to follow it up. This is a little too vague to be a f&f rule in Gigerenzer's rigorous sense, but it does have the virtue of calling attention to the distinction between making a decision and following up the decision made. The decision itself may be f&f, while the resulting research is likely to be anything but f&f.

We must also have a way of resolving conflicts among rules (including time-order conflicts, given that research time is itself a costly resource), which is why one might consider a production system of such rules, with some sort of prioritizing meta-rules. An example of a conflict is that the researcher may be presented with many choices. Satisficing prohibits trying to determine which choice has optimal heuristic potential. One possible rule is to pick that choice endorsed by the gut instincts of the key researchers.

Thus a heuristic consideration should not be eliminated simply because it does not contribute to a pre-established research goal. Part of the research exercise, after all, is to better identify and articulate fruitful goals and resulting problems. This is one area in which politically controlled funding agencies such as the U.S. National Science Foundation and the British Research Assessment Exercise (which imposes evaluation metrics on university research) run afoul of good research practice, by making it difficult to depart from the funded research proposal (Gillies 2008).

[14]It is no wonder that those positivists who drew the invidious context-of-discovery/context-of-justification distinction on traditional logical grounds found context of discovery to be non-logical and hence non-epistemic.

(o) It is also important not to restrict the application of f&f heuristics to individuals at the laboratory bench, for these heuristics are potentially applicable at all levels and with different degrees of precision as regards goals. Many "rules of thumb" for productively organizing research can probably be stated as f&f heuristics. For example, having regular afternoon teas is a British practice that brings researchers of different stripes together in an attempt to overcome narrow specialist silo or tunnel effects. Other simple rules of thumb govern the layout of lab space, informal meeting places in new buildings, institutional grant-processing procedures and the like that improve the general research culture by contributing to the economy of research.
Of course, the f&f decision at one level may result in a lot of work and cost at another. Here are three quick examples. The method of hypothesis may focus only on some known aspects of the phenomenon, namely those that it predicts, and disregard the others. And even the five-step "scientific method" taught in schools looks fairly fast and frugal if we formulate it at a high enough level of abstraction: "Formulate and test the hypothesis that would explain the puzzling phenomenon." However, such a rule is not computable, for it omits any instruction about how to find a hypothesis worth testing and how to devise a suitable test (Lenat 1978 again). Second, "Build the new laboratory complex with lots of space for informal communication." Third, the peer-review process used by journals is typically a f&f heuristic for filtering submissions: e.g., (i) Send article to three reviewers. (ii) If two of three agree on accept or reject, make the corresponding decision. (iii) If step (ii) fails, send submission to one more reviewer, etc.[15]

(p) Most of the examples discussed by the ABC Group concern empirical regularities. However, f&f heuristics might be used to make theoretically deeper moves. One is the aforementioned tools-to-theories heuristic, by which scientists attribute to the system whose behavior they are trying to understand the very same sort of cognitive tools the scientists themselves have developed to analyze such phenomena (Gigerenzer 1991; Gigerenzer and Sturm 2007). An example is the unconscious inference model of perception from Helmholtz to Richard Gregory and beyond, in which the cognitive system of the perceiving subject is hypothesized to compute visual attributes much as a team of scientists would. This move can be formulated as a one-step f&f rule: If methodological model M, constructed from the toolset T, adequately fits an entity's behavior, then attribute M to the organism or device itself as the real process that generates the behavior (Gigerenzer et al. 1999, 23). This will be possible only when the research methodology has matured, of course. Some historical uses of the heuristic are suspect, as Gigerenzer and Sturm and others

[15]The peer-review process used by most journals and funding agencies has recently come under fire. See, e.g., Braben (2004) and Gillies (2008).

have noted. However, the move may advance research at least by providing a "how possibly?" explanation of the behavior. I discuss the tools-to-theories heuristic in (Nickles, draft-2015). Given that the preparatory work that makes the tool-to-theories move possible is labor intensive and includes practical familiarity with the methods as well as theoretical work (but not yet attribution to the organism or artifact), should this work count against the heuristic's being f&f—or not, since it is work that is already done anyway?

(q) Another theoretical move is this. Much scientific research proceeds by modeling current problems on one or more previously solved problems, as Kuhn emphasized in his treatment of exemplars (1970a). He compared these scientific precedents to the use of precedents in legal reasoning. In *Structure* Kuhn limited the use of exemplars to normal science, so at first it would seem that Kuhn's exemplars face somewhat the same problems as Gigerenzer's f&f heuristics, namely, that they work well only for relatively crisp problems in domains in which a good deal of structure is already known. Unfortunately, Kuhn exaggerated the difference between normal and revolutionary science, as when he claimed that a revolution eliminates or replaces all of the old exemplars. However, he hinted in other passages that one can often trace a historical continuity across the revolutionary transition (1962, Chaps. X, XIII). For example, several special relativity and early quantum theoretic exemplars are modeled on those of classical physics. Nor (going still further beyond Kuhn) is such direct modeling of problems upon problems restricted to the domain in which the original exemplar was formulated. On the basis of either mathematical similarity or physical similarity, trans-domain modeling has often been fruitful, indeed sometimes the basis of major breakthroughs. The first order of business for many researchers is surely to locate precedents in their field or (going beyond Kuhn) in other fields that bear some similarity to their own research problems. Above I suggested a simple f&f rule for literature search. One difficulty to be faced in this attempt to merge Kuhn and Gigerenzer is that the latter treats f&f heuristics as explicit rules for decision-making, whereas Kuhn adamantly criticized a rules approach.[16]

To sum up: these bits of reasons and evidence for the importance of "less-can-be-more" research strategies are admittedly a scattered, ragtag bunch and usually only marginally f&f at best; but collectively they do amount to something. On the basis of reasons such as these it seems clear that "less can be more" often works, indeed, that it *must* sometimes work, since we have gloriously succeeded in solving the Meno problem and in getting more for less, thanks to broadly Darwinian variation-selection procedures.[17]

[16] *Structure*, Chap. V, but see also his 1960s attempts at computer modeling in Kuhn (1970a, b).

[17] Darwin himself saw a connection to the Meno problem as is evident from his notebook entries. See Desmond and Moore (1991, p. 263).

6 Some Reasons for Pessimism

I list here some challenges to extending the f&f approach to reasonably wild frontier contexts. Some items are surely genuine difficulties, others are open questions. The ABC group is aware of all of these issues in some form (see, e.g., Todd and Gigerenzer 2000), but more attention is needed to frontier contexts, if there is hope of extending the approach there. As before, the various items intersect and overlap in various ways, but I shall continue to number them separately for ease of reference. In some cases I briefly suggest a response.

(i) *The tamed frontiers problem*. The entire motivation for the extension to frontier contexts is flawed, for regarding frontier research as the application of f&f rules reduces it to routine, and that is precisely what frontier research is not. Again, Kuhn rebelled at a rules approach even for normal science (1962, Chap. V), but he had in mind explicit rules of the sort that would appear in scientific texts.
Brief response. No one is claiming that f&f rules cover everything. Also, weaker variations of f&f heuristics may be instructive even where there are no known, strict f&f rules. The main suggestion here is to replace truth with fallible, not even reliable, heuristic appraisal. Instead of requiring that a correct outcome of a decision be true, we require only that it have a reasonable chance of leading to fruitful and sustained research. Can it count as satisficing with an aspiration level so low? Cues can sometimes be ranked. Heuristic fertility is a fuzzy notion, but, as noted above, it is more accessible than is truth, both prospectively and retrospectively.

(ii) *The strong requirements problem*. The frugality of frontier research contexts does not yet establish the existence of usable f&f heuristics. Informationally frugal heuristics are not necessarily fast, and fast heuristics need not be frugal. In frontier contexts, imposing f&f or reliabilist requirements on heuristic thinking would throw out the baby with the bathwater.

(iii) *The evidence problem: descriptive*. What is missing from the ragtag list of Sect. 5 is hard evidence (e.g., historical cases or survey data) of the type that Gigerenzer and the ABC Group provide for the several successful one-, two-, or three-step heuristics that they discuss.

(iv) *The evidence problem: normative*. In frontier contexts there will usually be insufficient evidence available in advance to validate the heuristics. Hence, the heuristic approach can provide no guidance to research. How can researchers know that they are applying validated f&f rules for this domain in order to learn the domain structure, when validation presupposes that the structure is already known, or at least that there is already a successful record of application at the more superficial level of behavioral prediction? How can the ecological validity of f&f heuristics in frontier contexts be tested? What benchmarks could possibly be used in the case of fairly wild frontier research?
Response. One possible answer is that it cannot be known until domain structure is worked out well enough that certain kinds of problems now

become routine. If there are heuristic shortcuts in dealing with these problems that were used early the research, they can now be vindicated retrospectively. There are many instances in the history of science in which non-rigorous tricks used successfully (say for computation) early in the game were later vindicated or validated in some form, e.g., by showing that later, rigorous results justified the shortcut. Two examples would be the use of infinitesimals in early calculus and the early use of the Dirac delta function. (I do not here examine whether any of their uses were f&f.)

(v) In frontier contexts the goals, constraints, and hence problems, are often ill structured, making rigorous computability a severe (and still too traditional) a requirement to meet. (Cf. xii.)

(vi) *The multiple goals/tradeoffs problem*. Wise researchers retain multiple goals, with the result that decisions involve tradeoffs, whereas f&f heuristics are non-compensatory. There are also values tradeoffs, specific and general, fertility versus truth being a big, general one.

(vii) *The dilemma*. In sum, we face a dilemma. Insofar as we relax the standards for f&f heuristics in order to adapt the approach to frontier research, we lose the advantage of genuine f&f rules. But insofar as we enforce rigorously the requirements for f&f heuristics, we lose our purchase on frontier research.

(viii) *The ignorance-of-domain-structure problem (continued)*. The favorite ABC examples are not frontier examples (with the exception of tools-to-theories). In the case of the city-size decision problems, for example, we know that the f&f heuristic works, where it does, *because the analysts (Gigerenzer in this case) but not the subjects tested already possess the relevant information on the statistical structure of the domain* (sufficiently accurate information on city populations, in rank order). But that is precisely the sort of information that is typically lacking at research frontiers. Filling out the domain structure is the overall goal of the research, after all, and the researchers are in the position of unwitting subjects rather than knowledgeable analysts. At that point there do not yet exist knowledgeable analysts (in fact, none who will know the truth about nature with the assurance of city population statistics). This is why instruction in science courses leads to failure to understand the frontier problem, for the teachers and the textbook writers have the position of nearly omniscient analysts. One can even get a science degree without ever confronting frontier problems.

To suppose that domain structure is already known would beg the research question in somewhat the same way as did Simon's early programs that allegedly rediscovered Kepler's laws, the law of specific heats, Ohm's law, and others (Langley et al. 1987). Supposing that the key conceptual and empirical-organizational work has been done (identification of the key variables, the way to measure them, etc.) reduces frontier science to the routinized computations characteristic of textbook science—the eventual, pedagogical product of the frontier research. At genuine frontiers the ecological blade of Simon's scissors is largely missing. Despite his historical

sensitivity, Kuhn did something of the same thing in dealing successfully with context of discovery (insofar as he did) only by further taming the already non-wild frontier of normal science.[18]

(ix) *Confusion of analysts' and subjects' knowledge*. Another way of stating this difficulty is that the successful German guessers of American cities' relative populations do not themselves know that they have in hand a reliable heuristic. They are merely guessing. Only the analysts know, after the test results are in. (Gigerenzer himself was initially quite surprised at this result.) At the far frontier, everyone is merely a guesser, using something fairly close to trial and error. We must carefully distinguish the knowledge of the analysts from that of their subjects. A reliabilist take on the situation does not solve the problem either. Alvin Goldman holds that the successful Germans used what was in fact a reliable process and thus did know something about the domain, although they were not aware of that fact (commentary on Todd and Gigerenzer 2000). But if heuristics are to play an ameliorative or normative role at the frontier, don't they have to be known and applied explicitly? Well, perhaps the intuitions of experts are a counterexample, but are the successful German students experts? Surely not.

(x) *Where do heuristics come from*? To suppose that the heuristics are already known and available at the frontier is another way of framing the above challenges. Frontiers often call for the development of new methods or the modification of old ones. Even in the tools-to-theories case it would be misleading to say that method completely anticipates "theory," although it does anticipate the projection of the developed theory onto the subjects themselves. Those people who believe in "the scientific method" for doing successful science in all contexts run afoul of the aforementioned No Free Lunch Theorems. Moreover, they are strangely committed to a sort of intelligent design rather than an evolutionary conception of how new knowledge comes into existence (Nickles 2003).

(xi) *The ambiguity problem*. There seems to be an ambiguity in the term 'domain specific'. Gigerenzer's f&f rules are domain specific when it comes to successful application, but which methods are not? The question is whether they are domain-specific in incorporating domain knowledge into the rule or algorithm itself. To take a simple case, the heart attack assessment example at the beginning of *Simple Heuristics that Make Us Smart* incorporates domain-specific medical terms and threshold numerical values relevant to that domain. There is nothing like that in "take the best" and similar heuristics themselves. As pointed out in Sect. 3 on frontiers, what makes a method powerful is its including domain knowledge within it, and ditto for instrumentation containing internal processing. I don't disagree that it is the implicit domain specificity of "take the best" and the like that gives them their power, but I suspect that specificity in the strong sense can convey more

[18]See the especially the Goldman, Gorman, and Wang comments on Todd and Gigerenzer (2000).

power, as well as explicitly identifying the domain of application. By contrast with the explicitly specific rules, "take the best" and its cousins are domain-free methods already there waiting on the shelf to try out in various contexts. In this respect they after all resemble the general rules of Newell and Simon (1972), such as means-ends analysis and backward chaining.

Response. We must not overplay the difficulty, for it has turned out that the research in both the Meehl and Gigerenzer lines has generalized surprisingly widely. For many real-world systems we have learned that figuring out what the primary one or two variables are is often enough to get a pretty good grip on what is happening. The different f&f heuristics help us to classify problems into types or classes that can or cannot be handled in such-and-such a manner (the toolbox idea). Moreover, Gigerenzer et al. speak of "biases" built into f&f heuristics and agree with Newell and Simon when the latter write: "To the extent that the behavior is precisely what is called for by the situation, it will give us information about the task environment" (1972, p. 55). The success or failure of application conveys domain knowledge, making f&f heuristics useful probes, e.g., to determine whether one or two key variables are especially salient. Similarly, applying f&f heuristics together with the hypothetical use of the tools-to-theories heuristic to cognitive systems provides information about how they work.

(xii) *The rules attribution problem*. What sense does it make to attribute rules to non-linguistic animals or even to human cognitive systems, given that most cognitive processing occurs below the level of explicit conscious reasoning? Is the f&f heuristic approach committed to subconscious languages of thought of some kind? Connectionists will surely disagree. Insofar as they are right, it would appear that attribution of rules to subjects is an application of the tools-to-theories heuristic that is metaphorical rather than representationally realist. Dan Dennett's intentional systems approach comes to mind as a sort of intermediate position (Dennett 1987).

(xiii) *The knowledge pollution problem (reprise)*. Ironically, as research continues and domain knowledge becomes richer, it may happen that a heuristic that worked well enough at the crude stages of work becomes less effective, simply because the researchers now know too much. This 'Goldilocks' phenomenon is nicely illustrated by Gigerenzer's ongoing toy example of the city population comparison task. A German who is quite ignorant about the USA will do badly, but so will a German who knows a great deal. Finding the happy medium is just another task for the ongoing research, since it cannot be known in advance.

Response. Such a research trajectory has a happy outcome, after all. Many heuristics are valuable as ladders that can then be thrown away when no longer needed. After all, the methods, including heuristics and standards, also evolve, as the research proceeds. In all creative fields the standards imposed tend to increase in strength as the field advances.

(xiv) *The big switch problem*. Gigerenzer and company sometimes speak of their heuristics as simple 'building blocks', each one amounting to a kind of cognitive trick or short-cut that works well for a particular domain. A reader can be left with the impression that something like a highly modular conception of cognition is the goal here. Is the ABC approach on the path to massive modularity and its difficulties? One of these is the "big switch" problem. Insofar as organisms solve problems by means of problem-specific tricks, how do they decide what the problem is, in the first place, and hence the specific trick?

Response. Evolved cue responses largely answer this question and similarly for evolved tool use by human beings. Many species also have some degree of learning within their own lifetime sufficient to produce deeply ingrained habitual responses ("second nature") that mimic biological instincts. Thus no high-level cognitive switch as a decision mechanism is needed. Still, the Big Switch problem can arise in early frontier research insofar as a bunch of domain-neutral (as far as the investigators know) heuristics such as "take the best" are already on the shelf, waiting to be used.

Dennett speaks of the evolutionary-wired-in reasons for behavior as "free-floating rationales," since the reasons cannot be articulated by the actors, except, to some degree by us humans on the basis of evolutionary studies (1995, pp. 133, 232, 1996, p. 49, 2013, Chap. 40). We might extend the idea of free-floating rationales to habits of action acquired within actors' lifetimes insofar as they are unable to articulate that wisdom. Thus expert scientists working at frontiers can have free-floating rationales for some of the decisions they make. Another Dennett phrase applies here as well: "competence without comprehension," at least full comprehension (2013, Chap. 39). At the frontier the expert scientist is in a much better position to inquire than a layperson, even when both are ignorant of the structure of the domain at that point.

Another sort of big switch problem is how decisions are made when the subject has many different ways of achieving the goal. This is the Bernstein Problem, after the Russian, Nikolai Bernstein. Consider all the ways in which an octopus can reach for something (all the degrees of freedom in play). Other things being equal, the ABC Group advises to pick randomly. Fine, but what sort of computational realization would this require in an octopus?

(xv) *Rules versus intuitions*? A skeptic might wonder whether heuristic *rules* are what is wanted at the frontier, where hunches, intuitions, instincts, or 'gut feelings' (Gigerenzer 2007) of experts may be in play. And there are a variety of people from Wittgenstein to Heidegger and Hubert Dreyfus to Kuhn who raise serious objections to rules approaches (but see Stanley 2011; Thagard 2004, 2008). On the one hand, we do not want to treat all decisions in a Kantian manner (according to the stereotype), as products of explicit deliberation. On the other, many experts employ the term 'decision' in a very broad sense that applies to children, the hunches of experts, to animals, and

even to machines. Fortunately, the issue of what counts as a decision is not one that I can delve into here.

(xv) Gigerenzer and company tend to treat "gut feelings" as the same as, or based on, f&f heuristics. Other writers consider this a conflation of two different things. For example, the Meehl tradition makes a sharp distinction between intuitive decision-making and decisions based on statistically supported rules.[19]

7 Conclusion

The straightforward extension of the ABC paradigm to frontier contexts faces challenges. While there are surely some f&f rules of some sort already used by researchers, it is not clear how far they take us beyond practical wisdom already widely known. The whole matter needs further discussion, with many concrete applications. In my view a good strategy is to explore variations on the f&f idea, along the lines of some of the suggestions above. The two theoretically deeper heuristics mentioned above are the tools-to-theories move and adaptations of Kuhnian exemplars. Exemplars are like mini-toolboxes, and a (perhaps complex) version of the recognition heuristic is in play here.

Acknowledgement Thanks to Emiliano Ippoliti for organizing this stimulating conference, and thanks to him and to Fabio Sterpetti for their infinite patience and for work on the volume. Thanks also to Markus Kemmelmeier for a helpful comment.

References

Bishop, M., Trout, J.D.: Epistemology and the Psychology of Human Judgment. Oxford University Press, Oxford (2005)

Braben, D.W.: Pioneering Research: A Risk Worth Taking. Wiley, Hoboken, NJ (2004)

Campbell, D.T.: Evolutionary Epistemology. In: Schilpp, P.A. (ed.) The Philosophy of Karl R. Popper, 1, pp. 412–463. Open Court, Lasalle, IL (1974)

Damasio, A.: Descartes' Error. G.P. Putnam, New York (1994)

Dawes, R.: The robust beauty of improper linear models in decision making. Am. Psychol. **34**(7), 571–582 (1979)

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

Dennett, D.C.: Darwin's Dangerous Idea: Evolution and the Meanings of Life. Simon & Schuster, New York (1995)

Dennett, D.C.: Kinds of Minds: Toward an Understanding of Consciousness. Basic Books, New York (1996)

Dennett, D.C.: Darwin's "strange inversion of reasoning". PNAS **106**(Suppl 1), 10061–10065 (2009)

Dennett, D.C.: Intuition Pumps and Other Tools for Thinking. Norton, New York (2013)

[19]See Meehl (1954), Bishop and Trout (2005) and Trout (2009, Chap. 5, "Stat versus Gut").

Desmond, A., Moore, J.: Darwin. Warner, New York (1991)
Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1979)
Gigerenzer, G.: From tools to theories: a heuristic of discovery in cognitive psychology. Psychol. Rev. **98**, 254–267 (1991)
Gigerenzer, G.: Gut Feelings. Viking Penguin, New York (2007)
Gigerenzer, G., Sturm, T.: Tools = theories = data? On some circular dynamics in cognitive science. In: Ash, M., Sturm, T. (eds.) Psychology's Territories: Historical and Contemporary Perspectives from Different Disciplines. Lawrence Erlbaum, Mahwah, NJ (2007)
Gigerenzer, G., Todd, P.M.: ABC Research Group: Simple Heuristics That Make Us Smart. Oxford University Press, Oxford (1999)
Gigerenzer, G.: Rationality for Mortals. Oxford University Press, Oxford (2010)
Gigerenzer, G., Hertwig, R., Pachur, T. (eds.): Heuristics: The Foundations of Adaptive Behavior. Oxford University Press, Oxford (2011)
Gillies, D.: How Should Research Be Organised? College Publications, London (2008)
Kahneman, D., Slovic, P., Tversky, A.: Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge (1982)
Klein, G.: Sources of Power: How People Make Decisions. MIT Press, Cambridge, MA (1999)
Kuhn, T.S.: The Structure of Scientific Revolutions. University of Chicago Press, Chicago (1962)
Kuhn, T.S.: Postscript-1969. Addition to the Second Edition, of Kuhn (1962). University of Chicago Press, Chicago (1970a)
Kuhn, T.S.: Logic of discovery or psychology of research? In: Lakatos, I., Musgrave, A. (eds.), Criticism and the Growth of Knowledge, pp. 1–23. Cambridge University Press, Cambridge (1970b)
Langley, P., Simon, H.A., Bradshaw, G., Zytkow, J.: Scientific Discovery: Computational Explorations of the Creative Processes. MIT Press, Cambridge, MA (1987)
Lenat, D.: The ubiquity of discovery. Artif. Intell. **9**, 257–285 (1978)
Levins, R.: The Strategy of Model Building in Population Biology. Am. Sci. **54**(4), 421–431 (1966)
Meehl, P.E.: Clinical versus Statistical Prediction. University of Minnesota Press, Minneapolis (1954)
Miller, A.: Inconsistent Reasoning toward Consistent Theories. In: Meheus, J. (ed.) Inconsistency in Science, pp. 35–41. Kluwer Academic Publishers, Dordrecht, NL (2002)
Newell, A., Simon, H.A.: Human Problem Solving. Prentice-Hall, Englewood Cliffs, NJ (1972)
Nickles, T.: Evolutionary Models of Innovation and the Meno Problem. In: Shavinina, L. (ed.) International Handbook on Innovation, pp. 54–78. Elsevier Scientific Publications, Amsterdam (2003)
Nickles, T.: Heuristic Appraisal: Context of Discovery or Justification? In: Schickore, J., Steinle, F. (eds.) Revisiting Discovery and Justification: Historical and Philosophical Perspectives on the Context Distinction, pp. 159–182. Springer, Dordrecht (2006)
Nickles, T.: Heuristic Appraisal at the Frontier of Research. In: Ippoliti, E. (ed.) Heuristic Reasoning, pp. 57–87. Springer, Dordrecht (2015)
Nickles, T.: The Crowbar Model of Method: Reflections on the Tools-to-Theories Heuristic (draft-2015)
Nickles, T.: Prospective versus retrospective points of view in theory of inquiry: toward a Quasi-Kuhnian history of the future. In Beaney, M., et al. (eds.), Aspect Perception after Wittgenstein: Seeing-As and Novelty. Routledge, London (in press)
Norman, D.: Things that Make Us Smart: Defending Human Attributes in the Age of the Machine. Addison-Wesley, Boston (1993)
Popper, K.R.: Objective Knowledge: An Evolutionary Approach. Clarendon Press, Oxford (1972)
Pretz, J.E., Naples, A.J., Sternberg, R.J.: Recognizing, defining, and representing problems. In: Davidson, J.E., Sternberg, R.J. (eds.) The Psychology of Problem Solving, pp. 3–30. Cambridge University Press, Cambridge (2003)
Rorty, R.: Contingency, Irony, and Solidarity. Cambridge University Press, Cambridge (1989)

Schooler, L.J., Hertwig, R.: How forgetting aids heuristic decisions. Psychol. Rev. **112**, 610–628 (2005)
Simon, H.A.: Administrative Behavior. Macmillan, New York (1945)
Simon, H.A.: Models of Discovery. Reidel, Dordrecht (1977)
Simon, H.A.: What is an explanation of behavior? Psychol. Sci. **3**, 150–161 (1992)
Stanley, J.: Know How. Oxford University Press, Oxford (2011)
Thagard, P.: Rationality and science. In: Mele, A., Rawling, P. (eds.) The Oxford Handbook of Rationality, pp. 373–379. Oxford University Press, Oxford (2004)
Thagard, P.: Hot Thought. MIT Press, Cambridge, MA (2008)
Todd, P.M., Gigerenzer, G.: Précis of Simple Heuristics that Make Us Smart. Behav. Brain. Sci. **23**(5), 727–741 (2000)
Trout, J.D.: Why Empathy Matters: The Science and Psychology of Better Judgment. Penguin, New York (2009)
Vredeveldt, A., Hitch, G.J., Baddeley, A.D.: Eye closure helps memory by reducing cognitive load and enhancing visualisation. Mem. Cognit. **39**(7), 1253–1263 (2011)
Wimsatt, W.C.: Reductionistic research strategies and their biases in the units of selection controversy. In: Nickles, T. (ed.) Scientific Discovery: Case Studies, pp. 213–259. Reidel, Dordrecht (1980)
Wimsatt, W.C.: Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality. Harvard University Press, Cambridge, MA (2007)
Wolpert, D.H.: The lack of *a priori* distinctions between learning algorithms. Neural Comput. **8**(7), 1341–1390 (1996)
Wolpert, D.H., Macready, W.G.: No Free Lunch Theorems for Search. Technical Report SFI-TR-95-02-010. Santa Fe Institute (1995)

Scientific Realism, the Semantic View and Evolutionary Biology

Fabio Sterpetti

Abstract The semantic view of theories is normally considered to be an account of theories congenial to Scientific Realism. Recently, it has been argued that Ontic Structural Realism could be fruitfully applied, in combination with the semantic view, to some of the philosophical issues peculiarly related to biology. Given the central role that models have in the semantic view, and the relevance that mathematics has in the definition of the concept of model, the focus will be on population genetics, which is one of the most mathematized areas in biology. We will analyse some of the difficulties which arise when trying to use Ontic Structural Realism to account for evolutionary biology.

Keywords Scientific realism · Structural realism · Semantic view of theories · Evolutionary biology · Population genetics · Models

1 Introduction

Recently, Steven French (French 2014) has claimed that Ontic Structural Realism (OSR), a position normally held by philosophers interested in metaphysically accounting for physical theories, may be, in combination with the semantic view of theories, fruitfully adopted also to account for biological theories, especially population genetics. The present work is aimed at assessing whether this proposal hits the mark or not.

This paper will firstly briefly present the context in which OSR has been developed, and what is the main problem that this position has to face (Sect. 2); then, it will briefly present the semantic view and three of the main problems that this position has to face (Sect. 3); then, the paper will take into account two different possible responses to the main problem of structuralism, i.e. Psillos' and French's

F. Sterpetti (✉)
Department of Philosophy, Sapienza University of Rome, Rome, Italy
e-mail: fabio.sterpetti@uniroma1.it

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_4

responses, and it will try to underline the difficulties of each position (Sect. 4). Finally, the paper will focus on one of the examples given by French to illustrate his proposal of adopting structuralism in biology, i.e. Price's Equation, and it will try to spell out the difficulties of supporting French's claims on the metaphysical significance of Price's Equation for population genetics (Sect. 5).

2 Scientific Structural Realism

2.1 Scientific Realism

Scientific Realism (SR) can be briefly described as the claim that our best scientific theories are true. As Saatsi and Vickers state: "scientific realists seek to establish a link between theoretical truth and predictive success" (Saatsi and Vickers 2011, p. 29). SR is based on a two step strategy(Ellis 2009): (1) infer from the empirical success the truth of the scientific theories; (2) infer from the truth of the successful scientific theories the existence of those entities which appear in such theories. So, the claim that theories are able to refer to such existing entities is justified from their empirical success, while this very same ability explains their predictive empirical success.

2.2 Truth

The concept of truth is central for SR. For example, Giere states that: "virtually every characterization of scientific realism I have ever seen has been framed in terms of truth" (Giere 2005, p. 154).The most shared view of truth among the realists is that of truth as correspondence. For example, Sankey states that: "correspondence theories which treat truth as a relation between language and reality are the only theories of truth compatible with realism" (Sankey 2008, p. 17).

Given that the crucial element in order to claim for the truth of a theory is the confirmation of such theory, and that confirmation doesn't allow to discriminate between the different parts of the theory which has been confirmed, and that theories usually contain theoretical terms, i.e. terms which refer to some unobservables, realists believe in the existence of the theoretical terms postulated by the confirmed theory.

2.3 The No Miracle Argument

The main argument to support SR is the No Miracle Argument (NMA). Putnam formulated the NMA as follows: "The positive argument for realism is that it is the

only philosophy that does not make the success of science a miracle" (Putnam 1975, p. 73). The central idea of the NMA is that the truth of a scientific theory is the best, or the only scientifically acceptable, explanation of its empirical success. The problem is that, given the traditionally accepted realist view of truth, claiming that the success of a theory is due to its being true would imply that such theory should not be radically modified over time or ever considered false.

2.4 The Pessimistic Meta-Induction

But the history of science seems not to allow us to support such a claim. The Pessimistic Meta-Induction (PMI), firstly developed by Laudan (1981), can be briefly summarized as follows:

1. The historical record reveals that past theories which were successful turned out to have been false.
2. So, our present scientific theories, although successful, will probably turn out to be false.
3. Therefore, we should not believe our best present theories.[1]

To face the PMI different strategies have been developed by the realists. Many of them try to show that despite the theory shift, something is retained from one theory to another, and that it is just such 'something' that the realist is committed to.

2.5 Scientific Structural Realism

The most credited position in this realist line of reasoning is Scientific Structural Realism (SSR). French states that SSR has been developed exactly "to overcome the so-called Pessimistic Meta-Induction, which presents the realist with the problem of accommodating the apparent historical fact of often-dramatic ontological change in science" (French 2011, p. 165). Even if SSR does not rely on the NMA, it is normally considered to be able to support *the intuition* at the origin of the NMA, i.e. that there is a deep correlation between success and truth.[2] SSR is articulated in two main positions: Epistemic Structural Realism (ESR), which claims that we can be realist only about the mathematical structure of our theories (Worrall 1989), and OSR, which claims that structure is all there is (Ladyman

[1]Magnus (2010, p. 804).

[2]Cf. French and Ladyman (2003, p. 45): "structural realism is supposed to be realist enough to take account of the no-miracles argument".

1998).[3] So, what is thought not to change during the theory shift by SSR is the mathematical structure of the theories. For example, Sneed says that "structuralists see the mathematical structures associated with a theory to be much more 'essential' features of the theory than the claims it makes. The claims may change with the historical development of the theory, but the mathematical apparatus remains the same" (Sneed 1983, p. 351).

2.6 *Scientific Structural Realism and the Semantic View*

The focus on the mathematical structures of the theories makes clear why those who support SSR usually support the semantic view of theories (Chakravartty 2001), the view according to which a theory is the class of its models. For example, Ladyman states that the "'semantic' or 'model-theoretic' approach to theories, […], is particularly appropriate for the structural realist" (Ladyman 1998, p. 417), and suggests that "structural realists adopt Giere's account of theoretical commitment: to accept a theory means believing that the world is similar or isomorphic to one of its models" (Halvorson 2012, p. 185). Indeed, the way in which models are intended in the semantic view is the same in which models are intended in metamathematics. Thus, they are mathematical structures. To see this is particularly easy: the relation of isomorphism, which is claimed to hold among the models of a theory by the semanticists, is defined exactly in the same terms in which the relation of isomorphism is defined in model theory. Semanticists look at Tarski as the initiator of the semantic view of theories (da Costa and French 2003) and explicitly adopt the Tarskian concept of model in their view. For example, Suppes claims that "'the concept of model in the sense of Tarski may be used without distortion and as a fundamental concept' in scientific and mathematical disciplines," and that "'the meaning of the concept of model is the same in mathematics and the empirical sciences';" so, we can conclude that for him "the Tarski concept of a model is a common formal framework for analysis of various uses of models in science and mathematics" (Suppe 2000, pp. S110–S111).

2.7 *Ontic Structural Realism and Mathematics*

If the structuralist supports ESR, she is committed to the indispensable role of mathematics in accounting for theory change in a realist fashion. But if she supports OSR, she is also committed to the *existence* of the mathematical structures which figure in the

[3]See Frigg and Votsis (2011) for a survey on SSR. For a definition of structure, cf., e.g., Ibidem, p. 229: "A structure *S* consists of (a) a non-empty set *U* of objects, which form the domain of the structure, and (b) a non-empty indexed set *R* (i.e. an ordered list) of relations on *U*, where *R* can also contain one-place relations".

theory, and so she has to face the risk of let her position become a full-blood Pythagorean position. Indeed, "Pythagoreanism […] is the teaching that the ultimate 'natural kinds' in science are those of pure mathematics" (Steiner 1998, p. 60).

This risk that the supporters of OSR have to face, has been labeled by French the 'Collapse Problem' (French 2014): if the world is isomorphic to theories, and isomorphism can hold only between mathematical structures, then the world is a mathematical structure. Tegmark, for example, "explains the utility of mathematics for describing the physical world as a natural consequence of the fact that the latter is a mathematical structure, and we are simply uncovering this bit by bit. […]. In other words, our successful theories are not mathematics approximating physics, but mathematics approximating mathematics" (Tegmark 2008, p. 107). In fact, Tegmark argues, "the external reality is" not "*described by* mathematics, […] it *is* mathematics […]. This corresponds to the 'ontic' version of universal structural realism […]. We write *is* rather than *corresponds* to here, because if two structures are isomorphic, then there is no meaningful sense in which they are not one and the same. From the definition of a mathematical structure […], it follows that if there is an isomorphism between a mathematical structure and another structure […], then they are one and the same. If our external physical reality is isomorphic to a mathematical structure, it therefore fits the definition of being a mathematical structure" (Ibidem).

3 The Semantic View of Theories and Biology

3.1 The Semantic View and Evolutionary Biology

Since the eighties many authors have been supporting the semantic view of theories as the best account of evolutionary biology, and tried to elaborate a semanticist account of evolution (Lloyd 1984; Thompson 1983; Beatty 1980). This view has rapidly become the received view. The reasons for proposing and accepting the semantic view as the best account of biological theories were basically two: (1) the difficulties afflicting the traditional syntactic account of theories; (2) the more specific fact that, given that biology is normally considered to lack general laws from which starting to axiomatize an entire field of research (Beatty 1995), the semantic view seemed to be more adequate to meta-theoretically describe the biological theories, directly presenting a set of their models, instead of trying to axiomatize them.

In what follows, we will firstly describe some of the difficulties which afflict the semantic view in general, and then a specific difficulty related to the attempt of semantically representing the evolutionary processes.

3.2 Two Main Difficulties of the Semantic View

The semantic view of theories can be seen as composed by two parts: the first which equates theories and classes of models, the second which defines the relation between such models and the empirical world (Halvorson 2012). Both these parts have been challenged.

Halvorson focuses his criticisms on the first part, and shows that such "first component is a mistake; i.e., a class of models is not the correct mathematical component of a theory" (Halvorson 2012, p. 189), because "this view equates theories that are distinct, and it distinguishes theories that are equivalent" (Ibidem, p. 183). In fact, Halvorson shows that there is no good notion of isomorphism between classes of models, and so that the semantic account fails to provide a satisfactory account of the identity of theories.[4]

This is a big problem for the semanticists, because the possibility of clearly identifying a theory is considered essential in order to give a realist account of the theory change which can avoid the PMI. Indeed, as we have seen above, to avoid the PMI has been the principal motivation for the development of SSR. Suppe explicitly states that the semantic view "is inadequate if it cannot properly individuate theories. Theories undergo development. This has implications for theory individuation," because the semantic view "essentially treats theory development as progression of successive theories," and he adds that he considers "theory individuation issues as make-or-break for any account of theories" (Suppe 2000, pp. S108–S109). To sum up: there is a deep relation between OSR and the semantic view. The issue of theory individuation is considered to be crucial by the semanticists themselves to determine whether the semantic view is a tenable position or not, and Halvorson's work seriously treats exactly such crucial requisite of the semantic view.

[4]For technical details and examples, see (Halvorson 2012). Here the room suffices just to sketch Halvorson's argument in five points: (1) given that, according to the semantic view, a theory is a class of models, if we have two classes of models, $\mathcal{M}$ and $\mathcal{M}'$, under which conditions should we say that they represent the same theory? (2) Semanticists (e.g., Giere, Ladyman, Suppe, van Fraassen) have not offered any sort of explicit definition of the form: (X) $\mathcal{M}$ is the same theory as $\mathcal{M}'$ if and only if (iff)… (3) "Suppe's claim that 'the theories will be equivalent just in case we can prove a *representation theorem* showing that $\mathcal{M}$ and $\mathcal{M}'$ are *isomorphic* (structurally equivalent)' […] just pushes the question back one level—we must now ask what it means to say that two classes of models are 'isomorphic' or 'structurally equivalent'" (Halvorson 2012, p. 190). (4) He then considers the following three proposals for defining the notion of an isomorphism between $\mathcal{M}$ and $\mathcal{M}'$: (a) Equinumerous: $\mathcal{M}$ is the same theory as $\mathcal{M}'$ iff $\mathcal{M} \simeq \mathcal{M}'$; that is, there is a bijection $F : \mathcal{M} \to \mathcal{M}'$. (b) Pointwise isomorphism of models: $\mathcal{M}$ is the same theory as $\mathcal{M}'$, just in case there is a bijection $F : \mathcal{M} \to \mathcal{M}'$ such that each model $m \in \mathcal{M}$ is isomorphic to its paired model $F(m) \in \mathcal{M}'$. (c) Identity: $\mathcal{M}$ is the same theory as $\mathcal{M}'$, just in case $\mathcal{M} = \mathcal{M}'$. (5) Finally, he tests such proposals and shows how they all fail, and thus makes it clear that it is impossible to formulate good identity criteria for theories when they are considered as classes of models.

Chakravartty has challenged the second part of the semantic view: given that the semantic view of theories is deeply related to the concept of truth as correspondence, if the semantic view is language independent, it cannot satisfactorily account for a correspondence relation. The problem is the following: how is it possible to state that theories (i.e. classes of models) correspond to the world, if the required realist definition of such a correspondence, e.g. that of Tarski, is relative to (formal) languages, while theories are supposed to be non-linguistic entities by the semanticists? (French and Saatsi 2006). Thus, in order to qualify the semantic view as being able to satisfy the realist's claims, it seems necessary *to presuppose* a correspondence relation between the mathematical structure of theories and the structure of the world.

The problem is precisely how to account for such presupposition. To answer the question: "why an abstract structure such as a mathematical model can describe the non abstract physical world?" the realist's response usually "depicts nature as itself a relational structure in precisely the same way that a mathematical object is a structure. On this view, if the mathematical model represents reality, it does so in the sense that it is a *picture or copy* […] of the structure that is there" (van Fraassen 2008, p. 242). So, the realist seems to subscribe to a sort of substantial correspondence theory of truth and representation. But when she is pressed by the fact that there are different ways of mathematically representing the physical world, the realist cannot do better than "insists that there is an essentially unique *privileged* way of representing: 'carving nature at the joints'. There is an objective distinction 'in nature' between on the one hand arbitrary or gerrymandered and on the other hand *natural* divisions" (Ibidem, p. 244). How does she justify such assertion? She doesn't: "*It is a postulate*" (Ibidem).

3.3 Gildenhuys' Attack on the Semantic View

Recently, Gildenhuys has pointed out a difficulty which afflicts the semantic attempt to describe evolutionary biology, more precisely population genetics (Gildenhuys 2013). In fact, in such approach models are normally described as states in the phase space of the represented systems:

> The models picked out are mathematical models of the evolution of states of a given system […]. This selection is achieved by conceiving of the ideal system as capable of a certain set of *states*—these states are represented by elements of a certain mathematical space […]. The variables used in each mathematical model represent various measurable or potentially quantifiable physical magnitudes […] any particular configuration of values for these variables is a state of the system, the state space or 'phase space' being the collection of all possible configurations of the variables.[5]

[5]Lloyd (1984, p. 244).

According to Gildenhuys, this view is inadequate to describe population genetics, because even if philosophers have argued that, "in comparison to the rival syntactic approach to scientific theories," the semantic view "provides a superior framework for presenting population genetics […], none of these writers has specified the class of mathematical structures that constitutes population genetics or any of its formalisms" (Gildenhuys 2013, p. 274).

Details are not relevant here, what is worth noting is that Gildenhuys not only shows that a clear and complete definition of the 'phase space' of a biological system has never actually been given, but also that it would not be easy to give it for *mathematical* reasons. Indeed, he focuses on Lloyd's formulation, according to which there are "two main aspects to defining a model. First, the state space must be defined […]; second, coexistence laws, which describe the structure of the system, and laws of succession, which describe changes in its structure, must be defined" (Lloyd 1994, p. 19). Then, Gildenhuys underlines the challenge "posed by the existence of coefficients in population genetics whose values are set by functions" (Gildenhuys 2013, p. 281).

The problem is that if we try to construct the class of such functions, we get a mathematical indefinite object, and so we are not able to give the phase space we should instead construct to model population genetics in a semanticist fashion. Roughly, the difficulty lays in how to relate the causal relations among the individuals and all the possible different fitness functions deriving by their interactions. Indeed, "by means of functions that feature relative frequency terms as arguments, frequency dependent selection functions capture causal dependencies among the individuals" (Ibidem, p. 282). The problem is that when the character of the causal dependencies among population members varies, "the functions that set fitness values must vary with them. This variation is not merely variation in the values taken by a fixed set of arguments arranged in a fixed functional form. Causal relationships among individuals […] need not be linear […]. Equally, they may feature exogenous variables" (Ibidem, pp. 282–283). Now, to see "how fitness functions […] pose an obstacle to using" the semantic approach to describe population genetics, notice that "alternative fitness functions are inconsistent. Different fitness functions serve as alternative equations for setting the value of a single parameter. They cannot appear side by side, then, in a system of coexistence laws" (Ibidem, p. 284). Moreover, if we try to construct the class of those functions, "there are good reasons to believe that the class of such functions is in fact mathematically indefinite" (Ibidem). Thus, the "charge is that […] we cannot specify the class of fitness functions" (Ibidem, p. 285) and so that we cannot construct a phase space for population genetics systems.

3.4 Longo's View on Phase Space in Biology

Gildenhuys' criticism may be seen in relation to a wider criticism on the very possibility of giving a phase space when dealing with biological entities, a kind of

criticism developed by Giuseppe Longo. Longo identifies the peculiarity of biology with respect to physics exactly in "the mathematical un-predefinability of biological phase space" (Longo and Montévil 2014, p. 195). Indeed, "in contrast to existing physical theories, where phase spaces are pre-given, if one takes organisms and phenotypes as observables in biology, the intended phase spaces need to be analyzed as changing in unpredictable ways through evolution" (Ibidem, p. 189). The fact is that "random events, in biology, do not 'just' modify the (numerical) values of an observable in a pre-given phase space, like in physics […]. They modify the very phase space" (Ibidem, p. 199). Thus, "one major aspect of biological evolution is the continual change of the pertinent phase space and the unpredictability of these changes." (Ibidem, p. 187). We will not try to assess such criticisms. What is worth noting here is that Gildenhuys and Longo underline the difficulty of giving a definite phase space when dealing with evolutionary theories. If, as many supporters of the semantic approach to evolution maintain, giving the phase space is essential for giving an account of biological theories in accordance to the semantic approach, and the semantic view is the view adopted by the structural realists, then the structural realists should face this kind of difficulty if they want to account for evolutionary biology in structural terms. They may face this difficulty either by giving a definite phase space for the system they want to model, or by showing that giving a phase space is not essential for their approach.

4 Structural Realism and the Collapse Problem

4.1 Facing the Collapse Problem

The Collapse Problem described above (Sect. 2.7) may be better understood considering it from a more general perspective: the pressure that mathematical platonists are doing on scientific realists in order to let the realists accept Mathematical Platonism (MP). For example, Psillos states that philosophy of science "has been a battleground in which a key battle in the philosophy of mathematics is fought […] indispensability arguments capitalise on the strengths of scientific realism, and in particular of the no-miracles argument […], in order to suggest that a) the reality of mathematical entities […] follows from the truth of […] scientific theories; and b) there are good reasons to take certain theories to be true" (Psillos 2012, p. 63).

The scientific realists may respond to the platonist's pressure in two ways: (1) widening their ontology to accept abstract objects; (2) continuing to rely on causality and trying to avoid a direct commitment to the existence of abstract objects. The first option has been taken, among others, by Psillos. The second option has been taken, among others, by French. In what follows we will describe their approaches and the main difficulties that afflict them.

4.2 An Analysis of Psillos' Approach

Psillos' approach is representative of the attempt of 'moving beyond causation' that many realists are pursuing exactly to take into account abstract objects and non-causal (basically, mathematical) explanations (Rice 2013). Indeed, classically, SR supported the idea that scientific theories should be taken at face-value. But, "a literal reading of scientific theories implies commitment to a host of entities with a (to say the least) questionable ontic status: numbers, geometrical points, theoretical ideals, models and suchlike" (Psillos 2011a, pp. 5–6). The difficulty of conceiving the reality of the *abstracta* leads the realist to face what Psillos has called the 'Central Dilemma': "Either theories should be understood literally, but then they are false. Or, they should not be understood literally, but then realism itself is false (at least insofar as it implies that theories should be understood literally)" (Ibidem, p. 6). Indeed, if we commit ourselves to a *correspondence* view of truth, how could we avoid to take scientific theories to be understood *literally*? But, if we take a theory to be literally true, then we should believe in the existence of, e.g., numbers. The problem is that numbers are defined by platonists in such a way that they are outside the reach of science, at least to the extent that science is considered to be indispensably related to causality, as the majority of the realists seems still to think. Indeed, the so called 'Eleatic Principle', which can be stated as: "everything that is real makes some causal difference to how the world is" (Newstead et al. 2012, p. 89), has been considered to be able to discriminate what exists, referring to causation, by many realists since a long time.

Moreover, since, as we have seen, ever more realists have started supporting the semantic view of theories, the problem of how to conceive of the nature of models has become central for the realists. In fact, if following the semantic view, "theories are taken to be collections of models," and models are considered as abstract objects, then "theories traffic in abstract entities much more widely than is often assumed:" the claim that models are abstract entities "is meant to imply that (a) they are not concrete and (b) they are not causally efficacious. In this sense, models are like mathematical abstract entities" (Psillos 2011a, p. 4). Thus, models are abstract in the same way in which mathematical entities are claimed to be abstract by the mathematical platonists. Indeed, MP can be briefly described as the claim that mind-independent mathematical abstract entities exist, and abstract entities are normally understood as 'non-spatiotemporally located' and 'causally inert' (Balaguer 2009).

To make SR and MP compatible, Psillos adopts an 'explanatory criterion' to determine his realist ontology, explicitly claiming that "something is real if its positing plays an indispensable role in the explanation of well-founded phenomena" (Psillos 2011a, p. 15). He clearly underlines the distance between the explanatory criterion and causality: "This explanatory criterion should not be confused with a causal criterion. It is not a version of the so-called Eleatic principle" (Ibidem).

So, even non causal objects exist, and it seems that their existence could be supported relying *only* on explanatory considerations. The argument could run

something like this: we believe in the existence of what appears to be indispensable in our scientific explanations, abstract entities are indispensible in our best scientific explanations, then abstract objects exist.

But an explanation could have a great explanatory power and *nevertheless* be false. So, we could risk to infer the existence of some object which doesn't *really* (i.e. from a realist perspective) exist. So, how can Psillos deem the explanatoriness and indispensability of an object be reliable means to infer the existence of such indispensable objects? The problem lays in the ambiguity of the way in which Psillos describes the relation between explanations and theories, and between theories and the world.

What Psillos doesn't explicitly state is that an inference from the explanatoriness to the existence could be sound only if one has already accepted the two classical realist assumptions concerning the truth, i.e.: (1) that truth is correspondence to the world, and (2) that our best scientific theories are true because they have *empirical success*. So, Psillos' argument could be restated as follows: given that we infer the truth of the theories from their success, and that abstract entities are indispensible for the very reaching of such success; if truth is correspondence, the success of a theory is (best) explained by the existence of the objects such theory refers to, then abstract objects exist; a scientific explanation relies on our *already selected*, i.e. empirically confirmed, best scientific theories; then, if in such explanation an abstract object appears to be indispensable, we can safely commit ourselves to its existence.

In this way, the dangerous equation between explanatoriness and confirmation in order to define ontological matters has been neutralized. This has been possible thanks to the occultation of the link between *confirmation* and the set of the *already accepted* theories, and the insertion of the requirement that acceptable *explanations* from which deriving our ontology can only be drawn from such set of empirically confirmed theories. In such a way, our ontology may *appear* as based exclusively on the 'explanatory criterion', which Psillos explicitly says that should not be confused with a causal criterion, but such explanatory criterion rests nevertheless upon confirmation, i.e. *empirical* success, which is at its turn normally intended and explained in *causal* terms.

So, despite what Psillos explicitly asserts on explanatoriness, confirmation still plays a crucial role for the realists in order to determine the truth of a theory, and causality still plays a crucial role in order to account for the way in which confirmation is obtained and detected.[6] In fact, in another work (published in the same period) Psillos himself explicitly affirms that the "best explanation (of the instrumental reliability of scientific methodology) is this: the statements of the theory which assert the specific causal connections or mechanisms in virtue of which methods yield successful predictions are approximately true" (Psillos 2011b, p. 23). Thus, it is not easy for a realist to accept *abstracta* and give up causality.

[6]At least in the measure in which Psillos doesn't give a different account of how to consider a theory to be empirically confirmed.

4.3 The Collapse Problem and the Problem of Representation

Besides its relation with MP, it is also important to underline the connection between the Collapse Problem and the problem of scientific representation. Indeed, the way in which theories are related to the world is the crucial problem of any kind of realism. For example, Hughes states that "in what sense […] a physical theory *represent* the world?" is one of "the key question that philosophers of physics need to address" (Hughes 1997, p. 344).

For SSR, as we have seen, the problem is also related to the problem of the applicability of mathematics: since models are normally intended as not being interpretable as literally true, and models are generally conceived of as mathematical models, the problem of the relation between models and the world amounts to the problem of the relation between mathematics and the world.

This is where the problems described above relative to the semantic view meet the problem afflicting OSR. In fact, being able to solve the problems of the semantic view would imply to be able to distinguish the mathematical from the physical, i.e. to solve the Collapse Problem, because it would amount to be able to identify the *right* mathematical formulation of a phenomenon among the many which are possible, given that we would be able to state which is *the correct representational relation* between a mathematical structure and the world, and this could be possible only if the mathematical structures and the world do not coincide, i.e. if they are not the same thing, as Tegmark claims.

The problem is that there is a sort of dilemma here for the realist: either to account for the relation between models and the world she insists on isomorphism, but then she has to face the 'Collapse Problem'; or she has to specify *which kind* of representational relation holds between the scientific theories and the world. The risk in this case is that there is not a fully realist account of the representational relation.

In a nutshell, a relation of isomorphism is a symmetric dyadic *objective* relation, while that of representation is an asymmetric 'triadic' *intentional* relation (Suárez 2003). In other words, if we have to move beyond isomorphism we have to introduce a *subject* in our picture. For example, Giere argues for an 'agent-based' conception of representation, and describes it as composed by the following elements: "Agents (1) intend; (2) to use model, M; (3) to represent a part of the world, W; (4) for some purpose, P" (Giere 2010, p. 269). Giere also states that it is important to note "that this conception presupposes a notion of representation in general. I doubt that it is possible to give a non-circular (or reductive) account of representation in general" (Ibidem, p. 274). This means that we do not have a formal account of the notion of 'representation' which is 'objective' in the same way in which the formal account we have of the notion of isomorphism is 'objective'. Bas van Fraassen clearly states that, terminology aside, "a scientific model is a representation. So even if a scientific theory is a set of scientific models, and those literally are mathematical structures, it does not follow that the identity of a

theory can be defined in terms of the corresponding set of mathematical structures without reference to their representational function" (van Fraassen 2014, p. 278).

The fact is that 'representation' is a *semiotic* relation: someone describes something as something else. There is an intrinsic subjective element in such an account which is unpalatable for many realists. A representation is subject- and context-dependent, while the realists aim at truth, and normally conceive of truth as mind- and stance-independent. Thus, if we maintain that we *represent* some phenomena, the problem is how to claim that our representation is the *right* representation, i.e. that it corresponds to reality. In other words, a satisfying realist account should concern only the model and the world, and should not include any reference to the knowing subject, while the notion of 'representation' seems to intrinsically involve a reference to the subject.

French is aware of the difficulty of solving such problem, but his answer is clearly unsatisfying. Indeed he simply acknowledges that the fact that:

> the relationship between any formal representation and the physical systems that it represents cannot be captured in terms of the former only […] led to the accusation that the structuralist who relies on such representational devices cannot give an appropriate account of the relationship between representations and the world in terms of those very representations. My response is that all current forms of realism must face this accusation, not just OSR.[7]

This is clearly an unsatisfying answer, because the fact that to account for the relation between theories and the world is a problem for any kind of realism, does not diminish the relevance of such problem for OSR.

4.4 French's Approach to the Collapse Problem

French faces the above described difficulties claiming both that (1) the distinction between mathematics and the physical has to be accounted for in terms of causality: "how we can distinguish physical structure from mathematical structure […]. The obvious answer is in terms of causality, with the physical structure we are to be realists about understood as fundamentally causal" (French 2011, p. 166); and that (2) we can secure the fact that our representation of the world is the right one relying on a sort of NMA: "In the realist case, we will have only the 'no miracles' intuition to justify our claim that our theories represent the structure of the world" (Brading 2011, p. 57, fn 24).

These two claims can be reduced to one. In fact: how can we be sure that causality is a genuine feature of the world and not a feature of our model, as the anti-realists would suggest, and so that we can safely rely on causality to distinguish the physical from the mathematical? We can be sure that causality is a feature of the world only if we adopt a realist stance: "Ladyman follows Giere […], who states

[7]French (2014, p. 195, fn 7).

that 'the crucial dividing line between empiricism and realism' concerns the status of modality, and urges that representing the world as modal is essential to the […] realist" (Ibidem, p. 58). Thus, the way in which we ground the claim that a physical structure does not coincide with a mathematical structure is based on a sort of NMA exactly in the same way in which the claim that our representation of the world is the right representation is based on a sort of NMA. To better see this point, let's follow the argument given by French on a related issue:

> On what basis can we ascribe lawhood […] to the world? Here we need to articulate the ascription […] within the structuralist framework: first of all, there is the attribution of laws, as features of structure, to the world. […]. The structuralist can follow this line and maintain a form of the No Miracles Argument, concluding that the best explanation for the success of a given theory is that 'its' laws are 'out there' in the world, as features of the structure of the world.[8]

Thus, the way to claim for a realist stance on causality is relying on a form of NMA. But the main appeal of SSR was exactly due to its supposed ability in avoiding the PMI, i.e. in supporting realism without directly relying on the NMA, given that the NMA is vulnerable to the PMI.

The fact is that if we accept OSR, we have to face the Collapse Problem. But, if we ground our defence from the Collapse Problem on causality and we justify our confidence in the fact that causality is a feature of the world and not a feature of the model relying on a sort of NMA, we find ourselves in the very same position in which the classical realists were in confronting the PMI. In fact, if we rely on the NMA, then we cannot be sure that a feature belongs to the world and not to the model, because we cannot exclude that our theory is false or incomplete, and so that the feature in question does not really correspond to anything in the world. If there is not a complete correspondence between our theory and the world, we are not able to safely state whether a feature belongs to the world or not. On the contrary, if a perfect correspondence holds, we can safely claim that everything that is in the model corresponds to something in the world. In this case we could safely assume that causality is in the world. But if we adopt OSR, to state a *perfect correspondence* between a theory and the world would amount to state that there is a relation of isomorphism between the mathematical structure of that theory and the world. But a relation of isomorphism may hold only between two mathematical structures. Thus, the Collapse Problem would step back again, and we would not be able to distinguish the mathematical and the physical anymore. Thus, we have to conclude that we are not able to certainly state whether causality is a feature of the world or not, and so that there is not an easy solution to the Collapse Problem for the realist based on causality. Both Psillos' and French's approaches seem to be inadequate.

[8]French (2014, pp. 274–275).

5 Structural Realism and Biology

5.1 Ontic Structural Realism and Price's Equation

Let's now turn to French's proposal of adopting OSR in dealing with biology. French's attempt to articulate a biological form of OSR faces "the obvious problem of a comparative paucity of mathematized equations or laws by means of which we can identify and access the relevant structures" (French 2014, p. 329). But such paucity of mathematical structures does not affect "all biological fields—population genetics and theoretical ecology are the exceptions" (Ibidem, p. 329, fn 8). Thus, French focuses on Price's Equation, which is "sometimes presented as representing 'The Algebra of Evolution', and which one could take as characterizing a certain fundamental—if, perhaps, abstract—and 'high-level' feature of biological structure" (Ibidem, p. 338). Indeed, Price's Equation is a central result in population genetics, and can be written in the following form:

$$\Delta z = \mathrm{Cov}(w,\ z) + \mathrm{Ew}(\Delta z) \tag{1}$$

where: 'Δz' is the change in average value of a character from one generation to the next; 'Cov(w, z)' represents the covariance between fitness w and character (action of selection); and 'Ew(Δz)' represents the fitness weighted average of transmission bias. The equation "separates the change in average value of character into two components, one due to the action of selection, and the other due to the difference between offspring and parents" (Ibidem). According to French, there is a sense in which Price's Equation "offers a kind of 'meta-model' that represents the structure of selection in general […] this covariance equation is independent of objects, rests on no contingent biological assumptions, and can be understood as representing *the modal, relational structure of the evolutionary process*" (Ibidem, italics mine).

5.2 Fisher's Fundamental Theorem of Natural Selection

In order to assess the claim made by French relative to Price's Equation, let's consider a special case of Price's Equation which has been widely debated: Fisher's Fundamental Theorem of Natural Selection (FTNS).

In fact, Price's Equation tells us exactly how much of a character will exist in the population in the next period. If we let the character equal fitness itself, then we get Fisher's theorem:

$$\Delta \bar{w} = \mathrm{Var}_{\mathrm{add}}(g)/\bar{w} \tag{2}$$

which can be read as: the change in average fitness from one generation to another equals the additive genetic variance in the first generation, divided by mean fitness. The additive genetic variance, i.e. '$Var_{add}(g)$', measures the fitness variation in the population that is due to the additive, or independent, action of the genes. In other words, it measures any gene's effect on fitness which is independent of its genetic background. Indeed, according to this view of population genetics, it is possible to see the total 'Genetic Variance' as the sum of the 'Additive Genetic Variance' and the 'Non-additive Genetic Variance'.

Since its formulation, the FTNS has received different interpretations. This is due to the unclear formulation of the FTNS given by Fisher in his writings. In fact, Fisher describes the FTNS as follows: "the rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time" (Fisher 1930, p. 35).

This formulation of the FTNS induced many authors (and Fisher among them) to compare the FTNS to the second law of thermodynamics, according to which entropy, on average, can never decrease. In this interpretation, the FTNS is thought to be able to give a formal definition of the 'directionality of evolution', i.e. to give a proof of the fact that fitness, on average, will never decrease. Such a kind of result would have explained the course of evolution, the development of more and more complex forms, without any reference to any kind of 'design' or 'teleological explanation'.

The problem is that "it is *simply untrue that the average fitness of a population undergoing natural* selection never decreases, so the rate of change of average fitness *cannot always* be given by the additive genetic variance" (Okasha 2008, p. 328). Okasha clarifies that "Fisher was not talking about the rate of change of average fitness at all, but rather the *partial rate of change which results from* [the direct action of] *natural selection* altering gene frequencies in the population, *in a constant environment*" (Ibidem, p. 329).[9]

This means that it is more careful to say that according to the FTNS when natural selection is the only force in operation, average fitness can never decrease.

5.3 The Failure of the Analogy with Thermodynamics

The problem with this more careful interpretation of the FTNS is that it undermines the analogy between the FTNS and the second law of thermodynamics. Indeed, "by

[9]To understand Fisher's understanding of the FTNS, we have to accept Fisher's view of 'environment': any change in the average effects constitutes an 'environmental' change. On the constancy of environment, cf. Okasha (2008, p. 331): "For Fisher, constancy of environment from one generation to another meant constancy of the average effects of all the alleles in the population. Recall that an allele's average effect (on fitness) is the partial regression of organismic fitness on the number of copies of that allele." Cf., also, Ibidem, p. 324: "an allele's average effect may change across generations, for it will often depend on the population's genetic composition".

Fisher's lights, natural selection will almost never be the *only force in operation; for by causing gene frequencies to change, selection* almost always induces environmental change, which is itself a force affecting average fitness" (Ibidem, p. 344). In fact "the environment in Fisher's sense will *not remain fixed, for selection itself alters it*" (Ibidem, p. 347). Details are not relevant here, the basic idea is that for Fisher, when natural selection operates, this fact directly alters both the mean fitness $\bar{w}$, and the 'environment', which at its turn alters the mean fitness $\bar{w}$. Thus, if the FTNS holds only when natural selection is the only force to operate in a constant environment, and if when natural selection operates, the environment cannot remain constant, then we should conclude that the situation described by the FTNS can never obtain. Thus, the analogy between fitness and entropy seems to fail.

5.4 Different Interpretations of the Fundamental Theorem

The biological meaning of the FTNS is at least contentious. Price (1972) and Ewens (1989) state that the FTNS is mathematically correct, but that it does not have the biological significance that Fisher claimed for it. On the contrary, Edwards (1994) and Grafen (2003) are much more sympathetic to Fisher. We cannot develop this issue here. What we will analyse is whether French's proposal may be useful in dealing with this topic. In other words, can OSR help us in trying to determine the biological significance of the FTNS?

The fact is that both the main interpretations of the FTNS agree on the mathematical validity of Fisher's result. Thus, since structuralism seems to be mainly committed to the mathematical structures of a theory, and since the two main interpretations of the FTNS do not diverge on this issue, OSR seems *prima facie* unable to asses which interpretation of the FTNS is to be preferred. But French's formulation of OSR states that structures have to be interpreted in causal terms, in order to face the Collapse Problem. Thus, it seems that French's approach should be able to assess which interpretation of the FTNS we should prefer. Indeed, the issue of the interpretation of the FTNS is deeply related to the issue of the nature of natural selection, i.e. to the debate over the *causal* nature of natural selection.

5.5 The Fundamental Theorem and the Nature of Natural Selection

Let's restate the issue we are dealing with: French supports the claim that Price's Equation gives us the modal structure of the evolutionary process. The FTNS is a special case of Price's Equation. Thus, we can infer that French supports an interpretation of the FTNS as a significant and substantial result referring to natural selection. Since French claims also that structures have to be interpreted as causal in

order to avoid the Collapse Problem, and since, as we have seen, he seems to support the biological significance of the FTNS, we should infer that he thinks that the relevant mathematical structure in this context, i.e. the FTNS, can be interpreted in causal terms, i.e. that the FTNS can be interpreted as referring to some causal process.

Now, the problem is that the process to which the FTNS refers is that of natural selection, and that the causal nature of such process is harshly debated. It is not relevant to take side on this issue here. What is worth underling is that, contrary to French's proposal, those who deny the causal nature of natural selection explicitly refer to the FTNS and its mathematical formulation and significance to support their claim that such a kind of result cannot be interpreted as referring to causal processes.

5.6 *The Causal Nature of Natural Selection*

The causal nature of natural selection has recently been put under severe scrutiny.[10] This sort of "causal scepticism is motivated by the fact that most, if not all, principles of evolutionary theory—such as the Price equation or Fisher's fundamental theorem of natural selection—are expressed by purely statistical terms such as variances or covariances" (Otsuka 2014, p. 2). For example, Matthen and Ariew state that the reason "for reifying natural selection […] lies in a[n] […] analogy between equations of population genetics—such as Fisher's Theorem—and certain equations of physics." (Matthen and Ariew 2009, p. 208). But this analogy is not well founded, because, unlike the models that we find in physics, the descriptions of natural selection "rendered by population genetics models are in general neither predictive nor explanatory," since "population genetics models are, in general, noncausal models" (Glymour 2006, pp. 369, 383). Moreover, natural selection itself is not a genuine feature of the world, it is just "ontologically derivative on individual-level events such as births, deaths, and mutations" (Matthen and Ariew 2009, p. 216).

This view of the nature of natural selection seems to conflict with French's idea that the mathematical structures of a theory have to be intended as *causal* and that mathematical structures give us the *fundamental* structures of the world. Indeed, it would be difficult to accept that we should think that the FTNS tells us something about the deep (causal) structure of the *evolutionary* process, i.e. the core process of biology, *if* such theorem refers to something which is not only an 'ontologically derivative' non-causal concept, but which is also not intrinsically related to anything biological in nature. For example, Matthen and Ariew state that natural selection "is not even a biological phenomenon as such. It holds in any history in which the terms of the theory can be jointly interpreted in a way that accords with

[10]See Otsuka (2014) for a survey.

the abstract requirements of the theory" (Ibidem, p. 222). To illustrate this point they show how the FTNS may be equally well applied to something which is certainly not a biological entity or process:

> Suppose that you have two bank accounts, one yielding 5 % interest and the other yielding 3 %. One can treat units of money in each account as the members of a population, and the interest rate as an analogue of fitness. Provided that no money is transferred from one account to another, one can treat these 'fitness' values as heritable—that is, the fitness of any particular (non original) piece or unit of money can be ascribed to the 'reproductive' rate (i.e., interest) on preexisting units of money. Thus you would have, as between the monies resident in the two accounts, variation in heritable fitness. On this interpretation, Fisher's Fundamental Theorem of Natural Selection applies to your bank accounts: it predicts (correctly) that the average interest earned by the two bank accounts taken together will increase in proportion to the variance of interest rates earned by your money in the two accounts.[11]

As already noted, here the issue is not taking side on the dispute over the nature of natural selection, but underling how the difficulty of assessing such nature poses a challenge for French's proposal.

5.7 *Ontic Structural Realism and the Meaning of Price's Equation*

The fact is that French seems to acknowledge the abstract and not-intrinsically biological character of Price's Equation. For example, he states that "Price himself emphasized that his equation could be used to describe the selection of radio stations with the turning of a dial just as easily as it could to describe biological evolution" (French 2014, p. 338, fn 8). Even if we accept, for the sake of the argument, that there could be a way to make this abstractness compatible with the claim that Price's Equation gives us the deep structure of *biological* evolution, there remains a problem. The problem is that several authors support the idea that it is possible to give a causal interpretation of the FTNS (Sober 1984; Millstein 2006). Indeed, the causal interpretation of the evolutionary principles "shows adaptive evolution as a genuine causal process, where fitness and selection are both causes of evolution" (Otsuka 2014, p. 1). But French does not refer to such authors. He instead explicitly refers several times to Samir Okasha's works (French 2014, Chap. 12). The point is that Okasha *does not* adopt a straightforward causal interpretation of the fundamental results of population genetics. For example, he states that "Price's equation is statistical not causal" (Okasha 2006, p. 25). Even more explicitly, Okasha states that Price's Equation "is simply a mathematical tautology whose truth follows from the definition of the terms. Nothing is assumed about the nature of the 'entities', their mode of reproduction, the mechanisms of inheritance, the genetic basis of the character, or anything else" (Ibidem, p. 24).

[11]Matthen and Ariew (2009, p. 222).

Thus, the problem is that French has not defended at all the claim that natural selection has a causal nature. The only mathematical structure taken into account by French, i.e. Price's Equation, which is supposed to be able to give us the deep structure of the evolutionary process, gives raise (mainly) to two different and incompatible interpretations. But French's structural proposal could help us in solving such issue only if the dispute over the causal nature of natural selection would have been already settled in favor of the causalists, and French says nothing on this point.

6 Conclusions

To conclude, let's briefly sum up the difficulties that French's proposal of adopting OSR in biology has to face, in order to assess whether this proposal gives us some advantage in philosophically dealing with biology. French seems to accept the causal nature of natural selection and the idea that structures have to be understood as causal structures, but Price's Equation and other population genetics results, as the FTNS, are often interpreted as giving non-causal explanations, also by those authors to whom French himself refers in order to illustrate his proposal. In focusing on the mathematical abstract features of such equations, French seems to think that such structural characteristics are enough to give us a structural description of biology. But then there is nothing which can help us in avoiding the Collapse Problem according to the French's own strategy to face this problem: there is nothing in French's proposal on Price's Equation which suggests that (1) the causal nature of natural selection can be safely shown to be a feature of the world and that (2) such feature of the world is correctly reflected by the population genetics models. Only if (1) and (2) obtain, in fact, the abstract structure given by Price's Equation could be interpreted both as the fundamental structure of biology and a causal structure. On the contrary, we have shown that there are relevant difficulties in showing that both these conditions hold.

Moreover, we have to stress that French's proposal does not confront at all with the traditional semanticist approach to biology, which claims that it is necessary to give a *phase space* of the biological system we are modeling. French contents himself with referring only to some equations, and so he neither defends the possibility of giving a semanticist account without having to give the phase space, nor tries to give a complete phase space of the model he is considering. He explicitly states: "How are we to identify these structures that we are supposed to be realist about? The most obvious route is via the equations" (French 2011, p. 165). But, as we have seen, this is an insufficient response to the detailed remarks made by Gildenhuys on the difficulty of constructing a phase space when dealing with population genetics.

Thus, it seems reasonable to conclude that in the biological domain, OSR has to face the same challenges that it has to face in other domains, and that dealing with biological issues does not give to OSR any peculiar help in facing those challenges,

e.g. the Collapse Problem. At the same time, OSR seems not able to solve any peculiar philosophical issue related to population genetics, since, on the contrary, it is the solution of a debated philosophical issue related to biology, such that of the nature of the natural selection, which, if given, could represent a support to a structuralist approach to population genetics.

References

Balaguer, M.: Realism and anti-realism in mathematics. In: Gabbay, D., Thagard, P., Woods, J. (eds.) Handbook of the Philosophy of Science. Volume 4. Philosophy of Mathematics, pp. 117–151. Elsevier, Amsterdam (2009)

Beatty, J.: Optimal-design models and the strategy of model building in evolutionary biology. Philos. Sci. **47**, 532–561 (1980)

Beatty, J.: The evolutionary contingency thesis. In: Wolters, G., Lennox, J.G. (eds.) Concepts, Theories, and Rationality in the Biological Sciences, pp. 45–81. University of Pittsburgh Press, Pittsburgh (1995)

Brading, K.: Structuralist approaches to physics: objects, models and modality. In: Bokulich, A., Bokulich, P. (eds.) Scientific Structuralism, pp. 43–65. Springer, Dordrecht (2011)

Chakravartty, A.: The semantic or model-theoretic view of theories and scientific realism. Synthese **127**, 325–345 (2001)

da Costa, N.C.A., French, S.: Science and Partial Truth. A Unitary Approach to Models and Scientific Reasoning. Oxford University Press, Oxford (2003)

Edwards, A.W.F.: The fundamental theorem of natural selection. Biol. Rev. **69**, 443–474 (1994)

Ellis, B.: The Metaphysics of Scientific Realism. Acumen, Durham (2009)

Ewens, W.J.: An interpretation and proof of the fundamental theorem of natural selection. Theor. Popul. Biol. **36**, 167–180 (1989)

Fisher, R.A.: The Genetical Theory of Natural Selection. Clarendon Press, Oxford (1930)

French, S.: Shifting to structures in physics and biology: a prophylactic for promiscuous realism. Stud. Hist. Philos. Biol. Biomed. Sci. **42**, 164–173 (2011)

French, S.: The Structure of the World. Oxford University Press, Oxford (2014)

French, S., Ladyman, J.: Remodelling structural realism: quantum physics and the metaphysics of structure. Synthese **136**, 31–56 (2003)

French, S., Saatsi, J.: Realism about structure: the semantic view and nonlinguistic representations. Philos. Sci. **73**, 548–559 (2006)

Frigg, R., Votsis, I.: Everything you always wanted to know about structural realism but were afraid to ask. Eur. J. Philos. Sci. **1**, 227–276 (2011)

Giere, R.N.: Scientific realism: old and new problems. Erkenntnis **63**, 149–165 (2005)

Giere, R.N.: An agent-based conception of models and scientific representation. Synthese **172**, 269–281 (2010)

Gildenhuys, P.: Classical population genetics and the semantic approach to scientific theories. Synthese **190**, 273–291 (2013)

Glymour, B.: Wayward modeling: population genetics and natural selection. Philos. Sci. **73**, 369–389 (2006)

Grafen, A.: Fisher the evolutionary biologist. Statistician **52**, 319–329 (2003)

Halvorson, H.: What scientific theories could not be. Philos. Sci. **79**, 183–206 (2012)

Hughes, R.I.G.: Models, the brownian motion, and the disunities of physics. In: Earman, J., Norton, J. (eds.) The Cosmos of Science, pp. 325–347. University of Pittsburgh Press, Pittsburgh (1997)

Ladyman, J.: What is structural realism? Stud. Hist. Philos. Sci. **29**, 409–424 (1998)

Laudan, L.: A confutation of convergent realism. Philos. Sci. **48**, 19–49 (1981)

Lloyd, E.A.: A semantic approach to the structure of population genetics. Philos. Sci. **51**, 242–264 (1984)
Lloyd, E.A.: The Structure and Confirmation of Evolutionary Theory. Princeton University Press, Princeton (1994)
Longo, G., Montévil, M.: Perspectives on Organisms. Biological Time, Symmetries and Singularities. Springer, Berlin (2014)
Magnus, P.D.: Inductions, red herrings, and the best explanation for the mixed record of science. Br. J. Philos. Sci. **61**, 803–819 (2010)
Matthen, M., Ariew, A.: Selection and causation. Philos. Sci. **76**, 201–224 (2009)
Millstein, R.: Natural selection as a population-level causal process. Br. J. Philos. Sci. **57**, 627–653 (2006)
Newstead, A., Franklin, J.: Indispensability without platonism. In: Bird, A., Ellis, B., Sankey, H. (eds.) Properties, Powers and Structures. Issues in the Metaphysics of Realism, pp. 81–97. Routledge, New York (2012)
Okasha, S.: Evolution and the Levels of Selection. Oxford University Press, Oxford (2006)
Okasha, S.: Fisher's fundamental theorem of natural selection—a philosophical analysis. Br. J. Philos. Sci. **59**, 319–351 (2008)
Otsuka, J.: Causal foundations of evolutionary genetics. Br. J. Philos. Sci. (2014). doi:10.1093/bjps/axu039
Price, G.R.: Fisher's 'fundamental theorem' made clear. Ann. Hum. Genet. **36**, 129–140 (1972)
Psillos, S.: Living with the abstract: realism and models. Synthese **180**, 3–17 (2011a)
Psillos, S.: The scope and limits of the no miracles argument. In: Dieks, D., Gonzalez, W.J., Hartmann, S., Uebel, T., Weber, M. (eds.) Explanation, Prediction, and Confirmation, pp. 23–35. Springer, Dordrecht (2011b)
Psillos, S.: Anti-nominalistic scientific realism: a defense. In: Bird, A., Ellis, B., Sankey, H. (eds.) Properties, Powers and Structures. Issues in the Metaphysics of Realism, pp. 63–80. Routledge, New York (2012)
Putnam, H.: Mathematics, Matter and Method. Cambridge University Press, Cambridge (1975)
Rice, C.: Moving beyond causes: optimality models and scientific explanation. Noûs (2013). doi:10.1111/nous.12042
Saatsi, J., Vickers, P.: Miraculous success? Inconsistency and untruth in Kirchhoff's diffraction theory. Br. J. Philos. Sci. **62**, 29–46 (2011)
Sankey, H.: Scientific Realism and the Rationality of Science. Ashgate, Burlington (2008)
Sneed, J.: Structuralism and scientific realism. Erkenntnis **19**, 345–370 (1983)
Sober, E.: The Nature of Selection. University of Chicago Press, Chicago (1984)
Steiner, M.: The Applicability of Mathematics as a Philosophical Problem. Harvard University Press, Cambridge (MA) (1998)
Suárez, M.: Scientific representation: against similarity and isomorphism. Int. Stud. Philos. Sci. **17**, 225–244 (2003)
Suppe, F.: Understanding scientific theories: an assessment of developments, 1969–1998. Philos. Sci. **67**, S102–S115 (2000)
Tegmark, M.: The mathematical universe. Found. Phys. **38**, 101–150 (2008)
Thompson, P.: The structure of evolutionary theory: a semantic approach. Stud. Hist. Philos. Sci. **14**, 215–229 (1983)
van Fraassen, B.C.: Scientific Representation. Oxford University Press, Oxford (2008)
van Fraassen, B.C.: One or two gentle remarks about Hans Halvorson's critique of the semantic view. Philos. Sci. **81**, 276–283 (2014)
Worrall, J.: Structural realism: the best of both worlds? Dialectica **43**, 99–124 (1989)

Models of the Skies

Emily Grosholz

Abstract Reasoning that adds content to scientific theories typically moves between the task of accurate reference and taxonomy, and the task of analysis, the theoretical search for conditions of intelligibility. Here we examine the development of models of astronomical systems, beginning with the early 17th century models of the solar system, and ending with late 20th century models of galaxies. In each case, we note both the distinction, and the interaction, between the aims of reference and analysis, and the ways in which disparate modes of representation combine to enlarge scientific knowledge.

Keywords Reference · Analysis · Scientific model · Astronomy · Cosmology · Kepler · Newton · Laplace · Clausius · Herschel · Rosse · Hubble · Zwicky · Rubin

1 Introduction

1.1 Ampliative Reasoning

Philosophers of mathematics and science have increasingly insisted on the importance of understanding ampliative reasoning, reasoning that adds content and yet is "rational" in a sense that goes beyond deductive and inductive logic. The writings of Carlo Cellucci and Nancy Cartwright come immediately to mind. We understand such reasoning variously, sometimes as a search for the solution of problems and sometimes as a search for the conditions of intelligibility of problematic things. Thus we are interested in the ampliative thrust of certain methods, notations and imaging, the inexhaustibly enigmatic nature of mathematical and

E. Grosholz (✉)
Department of Philosophy and Center for Fundamental Theory,
The Pennsylvania State University, University Park, USA
e-mail: erg2@psu.edu

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_5

natural objects and systems, and the important role played in research by the conjunction of different modes of representation, including iconic images, diagrams, and two- or three-dimensional displays alongside formal languages and natural language. The primacy of the closed, homogeneous systems of deductive logic (so limited in their ability to refer) for understanding how we reason is thus challenged, and so too are assumptions about how we integrate mathematics, physical science and empirical data. One of the central concerns is the nature of 'models' and how they bring the natural world and mathematics into rational relation.

1.2 Reference and Analysis

I begin by observing that productive scientific and mathematical discourse must carry out two distinct tasks in tandem: an analysis or search for conditions of intelligibility or solvability, and a strategy for achieving successful reference, the clear and public indication of what we are talking about, which often involves a search for the conditions of intelligibility of problematic objects. Difficulties often arise for mathematicians and scientists because modes of representation apt for analysis may prove to be inapt for successful reference, and vice versa. Sometimes the task of analysis is more difficult, and lags behind; sometimes the task of reference is more difficult; improvements in reference may lead to improvements in analysis, and vice versa, but often the different tasks, and attendant modes of representation, are hard to reconcile. One of the most effective ways to bring mathematics and the world into rational relation is to combine referential and analytic discourses, a task which inevitably involves the construction of successful models.

In his essay "Mathematics, Representation and Molecular Structure," Robin Hendry notes that Nancy Cartwright (and Margaret Morrison) distinguish strongly between two kinds of models (Hendry 2001). On the one hand, philosophers like Bas van Fraassen pay most attention to theoretical models, which as in model theory are structures that satisfy a set of sentences in a formal language: such structures are themselves organized as a language, so that the sentences of the formal language are true when interpreted in terms of the object-language. On the other hand, philosophers like Cartwright and Morrison remind us of the importance of representational models, where the relevant relation is not satisfaction (as between a meta-language and an object-language), but representation (as between a discursive entity and a thing that exists independent of discourse), like the iconic images that represent molecules. (This one is benzene, C_6H_6.) (Fig. 1).

Different models, or modes of discourse bring out different aspects of the 'target system.' Those that help us to elaborate theory and the abstract, highly theoretical networks that lead to scientific explanation, typically differ from those that help us to denote, to single out the intended target system. The relation between metatheory and object theory is isomorphism; but isomorphism leaves us adrift in a plurality of

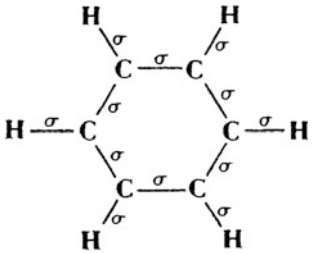


Fig. 1 Benzene molecule

possible structures. And scientists cannot allow themselves to drift in that way: Hendry writes, "… we note that equations are offered not in isolation, but in conjunction with text or speech. This linguistic context is what determines their denotation and serves to make representation a determinate, non-stipulative relation that may admit of (degrees of) non-trivial success and failure. Natural languages like English, French or German equip their speakers with abilities to refer to their surroundings, and we can understand how equations can represent if they borrow reference from this linguistic context." (Hendry 2001). In sum, theoretical models are too general; they cannot help us refer properly to the things and systems we are trying to investigate. And referential models are too limited; they cannot offer the explanatory depth that theory provides.

1.3 Models

My intention in this essay is to show how models of the solar system, our galaxy and closest galaxy-neighbor Andromeda, and our cosmos have historically proved to be composites: in order to be effective, they must combine discursive and representational modelling in an uneasy but fruitful unity. We need a thoughtful account of the variety of strategies that scientists use to construct such composite models. The relative stability of successful models makes scientific theorizing possible; and the residual instability (which no logician can erase) leaves open the possibility of further analysis and more accurate and precise reference. Models are revisable not only because they are 'approximations' that leave out information, but also because they must combine both reference and analysis. To inquire into the conditions of intelligibility of formal or natural things, we may decompose them in various ways, asking what components they have and how those components are related, or asking what attributes go into their complex concepts. We can also ask what laws they satisfy. But in order to refer properly to a thing or system, we have to grasp it first as what it is, a whole relatively stable in time and space, its unity governing the complex structure that helps to characterize it.

Things and systems—both natural and formal—have symmetries and (since periodicity is symmetry in time) so do natural processes! Carbon molecules as they throb, snowflakes as they form, and solar systems as they rotate exhibit symmetries

and periodicities that are key to understanding what they are and how they work. Thus the shape (in space and time) of a system or thing is not, as Aristotle once claimed, a merely accidental feature. On the contrary, symmetry and periodicity are a kind of generalization of identity; they are the hallmark of stable existence. Symbolic modes of representation seem to be most useful for abstract analysis, and iconic modes of representation for reference: a representation of shape is often an important vehicle for referring. This is an over-simplification, however; tabulated data and data displayed to exhibit (for example) linear correlations have both symbolic and iconic dimensions, and most icons come equipped with indices that relate them to symbolic notation. Thus we should expect models to be both symbolic and iconic. And then it is rewarding to ask, how do those modes of representation interact, on the page and in thought?

2 Early Modern Astronomy

2.1 Tycho Brahe and Kepler

In the late sixteenth century and throughout the seventeenth century, the problem of reference in astronomy is important but less pressing than problems of analysis. The latter include debate over whether the sun or the earth occupies the center of the cosmos, and whether heavenly bodies move in circles at a constant speed or not. However, the objects in question are clearly defined: we stand on the earth, the sun and the moon are large, brilliant objects in the sky, and the planets are salient and distinctive in their movements. To refer, in one sense, all we have to do is point. But the very act of pointing out an item in the solar system is a tracking: such items move, so the question of how to characterize that movement must also arise. Tracking the objects of the solar system required, in the sixteenth century, a compass and a sextant or quadrant; Tycho Brahe used these instruments in an unusually consistent and careful fashion, calibrating his instruments regularly and measuring their positions at small temporal intervals all along a given orbit with unprecedented accuracy. Brahe's tables of planetary motion, the *Rudolphine Tables*, meant to supplant the 13th c. *Alphonsine Tables*, were published a quarter century after his death in 1627, by his collaborator Kepler. The *Tables* are the model, along with Kepler's ellipse: what do I mean by this claim? (Kuhn 1957).

Perusing the pages of the *Rudolphine Tables*, we see that the issues of reference and analysis cannot be thoroughly disentangled. First, from the way Kepler sets up the *Tables*, it is clear that he is using a heliocentric system with elliptical planetary orbits. This is noteworthy because Tycho remained opposed to the heliocentric hypothesis till the end of his life, and he died before Kepler worked out his laws of motion: the claim that the orbit of Mars is elliptical is first published in Kepler's *Astronomia Nova* (1609). Thus the tables embody and display two theoretical challenges to Aristotelian/Ptolemaic astronomy which Tycho himself never made.

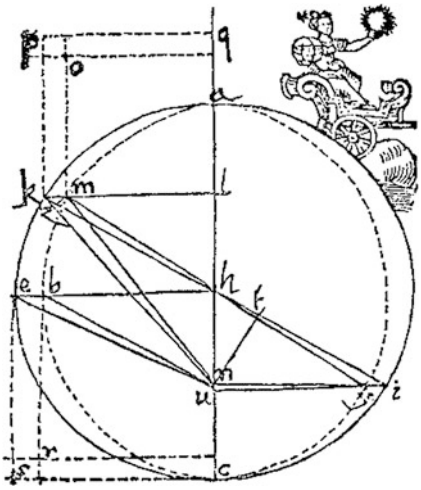


Fig. 2 Frontispiece, Astronomia Nova

Second, it was the very accuracy of Tycho's data that persuaded Kepler finally to abandon his devotion to the circle, and to search for other simple mathematical forms, at last settling on the ellipse. Unprecedentedly accurate and frequent tracking forced a change in conceptualization. Of course, it was also the highly theoretical mathematics of Euclid and Apollonius (newly available in the Renaissance) that offered a repertoire of forms to Kepler. His famous ellipse from the *Astronomia Nova* is given below (Kuhn 1957) (Fig. 2).

2.2 Galileo

As is well known, Galileo pounced upon the refracting telescope almost as soon as it was invented, made improvements to it, and turned it on the heavens. (Kepler was an enthusiastic supporter of Galileo's *Sidereus Nuncius* (1610), and himself used the telescope to look at the moons of Jupiter and the surface of the moon.) Sixty years later, Newton built the first reflecting telescope, using a concave primary mirror and a flat diagonal secondary mirror; this invention impressed both Barrow and Huygens, and led to Newton's induction into the Royal Society. From then on, improvements in our ability to refer to the objects of astronomy have depended on improvements in the material composition, size, and placement of telescopes. Galaxies and galaxy clusters, if they are not simply invisible, are at first mere smudges on the night sky (a few are visible as smudges to the naked eye). Either they are not recorded, or they are noted as 'nebulae,' clouds whose structure is just barely visible in the nineteenth century and whose composition remains mysterious until well into the twentieth. Like clouds, they seem to have no determinate shape; the discernment of galactic shape plays an important role in the development of twentieth century astronomy and cosmology (Wilson et al. 2014).

3 Newton's Principia

3.1 Principia, Book I

In Book I of Newton's *Principia*, Kepler's Second Law (that planets sweep out equal sectors in equal times: they accelerate as they get closer to the sun and decelerate as they get further away) is proved in Newtonian fashion in Proposition I, and Kepler's ellipse is the centerpiece of the diagram that accompanies the proof of the inverse square law, Proposition XI. This model is a geometric shape. In this case, however, what is modeled is only a fragment of the solar system, which consists of two bodies, the sun and a single planet. One might suppose then that the appropriate geometric model would simply be two points (. .); however, a quick look at Fig. 3 discredits that idea. For this model of the sun and a planet to exhibit the integrity of a system ("the System of the World") and to serve as the basis for building back more of the complexity of the known system with its sun, six planets and various moons in later models, two points side by side will not suffice. Rather, the model must include the spatial and temporal symmetries involved in the motions of the two bodies, and thereby, as we learn from Proposition XI, it must also express the dynamical nature of the interaction between them.

Proposition XI demonstrates that if a body in orbit around a center of force traces out an ellipse, then the force must obey an inverse square law. The geometrical array that we see in Figure I represents a planet at P in elliptical orbit around the sun at S, located at one of the two foci of the ellipse. Note that significant points, and therefore certain line segments and areas they delimit, on the geometrical construction (their significance is both geometrically and physically motivated) are labeled by letters, and that this array is surrounded by prose in Latin as well as proportions and equations involving the letters of those lettered points. The ellipse therefore appears as a palimpsest. It is at the same time a Euclidean-Apollonian mathematical object, with one set of internal articulations useful for discovering its mathematical properties; a tracking device for Kepler as he finishes compiling the Tables with Tycho's compass and sextant or quadrant, and therefore just an outline, since a trajectory is just a (projected) line across the sky; and finally as well Newton's construction, with a superimposed set of articulations for displaying temporal and physical properties. That final layer turns the array into the representation of a dynamical system, as the center of force is shown to obey the inverse square law. The ellipse thus becomes a model where the demands of reference and the demands of theorization are, in Book I, Proposition XI, happily reconciled. All the same, the multiple roles the ellipse is forced to play there in a sense destabilize the geometry and will ultimately lead to its re-expression in the Leibnizian form of a differential equation (Newton 1966; Grosholz 2007).

OTU
IRUM

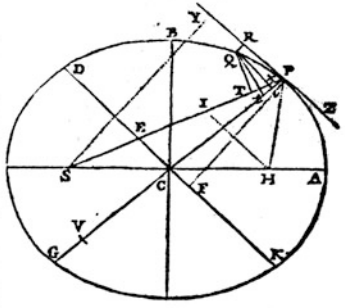
SECTIO III.

De motu Corporum in Conicis Sectionibus excentricis.

PROPOSITIO XI. PROBLEMA VI.

Revolvatur corpus in Ellipſi: requiritur Lex vis centripetæ tendentis ad umbilicum Ellipſeos.

Eſto Ellipſeos umbilicus S. Agatur SP ſecans Ellipſeos tum diametrum DK in E, tum ordinatim applicatam Qv in x, & compleatur parallelogrammum $Qx\,P\,R$. Patet EP æqualem eſſe ſemiaxi majori AC, eo quod acta ab altero Ellipſeos umbilico H linea HI ipſi EC parallela, (ob æquales CS, CH) æquentur ES, EI, adeo ut EP ſemiſumma ſit ipſarum PS, PI, id eſt (ob parallelas HI, PR & angulos æquales IPR, HPZ) ipſarum PS, PH, quæ conjunctim axem totum $2AC$ adæquant. Ad SP demittatur perpendicularis QT, & Ellipſeos latere recto principali (ſeu $\frac{2BC\,quad.}{AC}$) dicto L, erit $L \times QR$ ad $L \times Pv$ ut QR ad Pv, id eſt ut PE ſeu AC ad PC; & $L \times Pv$ ad $Gv\,P$ ut L ad Gv; & $Gv\,P$ ad $Qv\,quad.$ ut $PC\,quad.$ ad $CD\,quad.$; & (per Corol. 2 Lem. VII.) $Qv\,quad.$ ad $Qx\,quad.$ punctis, Q & P coeuntibus, eſt ratio æqualitatis; & $Qx\,quad.$ ſeu $Qv\,quad.$ eſt ad $QT\,quad.$ ut $EP\,quad.$ ad $PF\,quad.$, id eſt ut $CA\,quad.$ ad $PF\,quad.$ ſive (per Lem. XII.) ut $CD\,quad.$ ad $CB\,quad.$ Et conjunctis his omnibus rationibus, $L \times QR$ fit ad $QT\,quad.$ ut $AC \times L \times PCq. \times CDq.$ ſeu $2CBq. \times PCq. \times CDq.$ ad $PC \times Gv \times CDq. \times CBq.$ ſive ut $2PC$ ad Gv.



Sed

Sed, punctis Q & P coeuntibus, æquantur $2PC$ & Gv. Ergo & his proportionalia $L \times QR$ & $QT\,quad.$ æquantur. Ducantur hæc æqualia in $\frac{SPq}{QR}$ & fiet $L \times SPq.$ æquale $\frac{SPq. \times QTq.}{QR}$. Ergo (per Corol. 1 & 5 Prop. VI.) vis centripeta reciproce eſt ut $L \times SPq.$ id eſt, reciproce in ratione duplicata diſtantiæ SP. *Q.E.I.* LIBER PRIMUS.

Idem aliter.

Cum vis ad centrum Ellipſeos tendens qua corpus P in Ellipſi illa revolvi poteſt, ſit (per Corol. 1 Prop. X) ut CP diſtantia corporis ab Ellipſeos centro C; ducatur CE parallela Ellipſeos tangenti PR: & vis qua corpus idem P, circum aliud quodvis Ellipſeos punctum S revolvi poteſt, ſi CE & PS concurrant in E, erit ut $\frac{PE\,cub.}{SPq}$ (per Corol. 3. Prop. VII,) hoc eſt, ſi punctum S ſit umbilicus Ellipſeos, adeoque PE detur, ut SPq reciproce. *Q.E.I.*

Eadem brevitate qua traduximus Problema quintum ad Parabolam, & Hyperbolam, liceret idem hic facere: verum ob dignitatem Problematis & uſum ejus in ſequentibus, non pigebit caſus ceteros demonſtratione confirmare.

PROPOSITIO XII. PROBLEMA VII.

Moveatur corpus in Hyperbola: requiritur Lex vis centripetæ tendentis ad umbilicum figuræ.

Sunto CA, CB ſemi-axes Hyperbolæ; PG, KD diametri conjugatæ; PF, Qt perpendicula ad diametros; & Qv ordinatim applicata ad diametrum GP. Agatur SP ſecans cum diametrum DK in E, tum ordinatim applicatam Qv in x, & compleatur parallelogrammum $QRPx$. Patet EP æqualem eſſe ſemiaxi tranſverſo AC, eo quod, acta ab altero Hyperbolæ umbilico H linea HI ipſi EC parallela, ob æquales CS, CH, æquentur ES, EI; adeo ut EP ſemidifferentia ſit ipſarum PS, PI, id eſt (ob parallelas IH, PR & angulos æquales IPR, HPZ) ipſarum PS, PH, quarum differentia axem totum $2AC$ adæquat. Ad SP demittatur perpendicularis QT. Et Hyperbolæ latere recto principali (ſeu $\frac{2BCq}{AC}$) dicto L, erit $L \times QR$ ad $L \times Pv$ ut QR ad Pv, id eſt, ut PE ſeu AC ad PC; Et $L \times Pv$ ad $Gv\,P$ ut L ad Gv;

G

Fig. 3 Newton, *Principia*, book I, section III, proposition XI

3.2 *Principia, Book III*

In Book III of the *Principia*, Newton elaborates his theory and enriches his model, building in further complexity, to show that he can account for further tabular evidence compiled by other astronomers around Europe. He accounts for perturbations in the orbit of the moon in terms of the gravitational pull of both the earth and the sun, and goes on to account for the tides; he explains the orbits of comets as they intermittently visit the solar system; and he generalizes the law of universal gravitation. The problems left for the next generation by Newton's Book III are therefore, in his opinion, puzzles of normal science. (Kuhn apparently concurs in Newton's assessment.) On this account of scientific progress, the puzzles of reference are to locate and measure the movements of more and more astronomical items, and so to make sure that they accord with Newton's three laws of motion and the law of universal gravitation. Existing theory, expressed in the formal (highly geometrical) idioms of the *Principia*, will cover and explain observation, and prove adequate to solving the puzzles of theory, which include first and foremost how to move from the 2-body problem to the 3-body problem to the *n*-body problem.

Newton's Law of Universal Gravitation states that, in the case of two bodies, the force acting on each body is directly proportional to the product of the masses, and inversely proportional to the square of the distance between their centers; and it acts along the straight line which joins them. He also shows that gravity acts on the bodies, when they are solid spheres, in just the same way that it would act on point particles having the same masses as the spheres and located at their centers. This allowed the formulation of the *n*-body problem, which models a group of heavenly bodies: consider n point masses in three-dimensional Euclidean space, whose initial positions and velocities are specified at a time t_0, and suppose the force of attraction obeys Newton's Law of Universal Gravitation: how will the system evolve? This means we have to find a global solution of the initial value problem for the differential equation describing the *n*-body problem (Diacu and Holmes 1996).

But here is the irony: the differential equations of the two-body problem are easy to solve. (Newton's difficulties with his own much more geometric formulation in Book I, Propositions XXXIX–XLI indicate the superiority of the idiom of differential equations here. At this point in the *Principia*, he really needed Leibniz's help.) However, for n larger than 2, no other case has been solved completely. One might have thought that "reducing" the models to differential equations would have made the solution of these centrally important problems about the solar system straightforward. But on the contrary, the equations articulated the complexity of the high-dimensional phase spaces needed to express accurately the physical situation (sub-systems of the solar system), as well as the great difficulty of finding complete solutions. The severe difficulty of the *n*-body problem drove the development of physics for many decades. The work of Leibniz, Euler, Lagrange, Laplace and Hamilton replaced Newton's Laws with a single postulate, the Variational Principle, and replaced Newton's vectorial mechanics with an analysis in which the fundamental quantities are scalars rather than vectors, and the dynamical relations

are arrived at by a systematic process of differentiation. Lagrange's *Méchanique Analytique* (1788) introduced the Lagrangian form of the differential equations of motion for a system with n degrees of freedom and generalized coordinates q_i ($i = 1, \ldots n$). This re-writing of the equations allowed physicists to choose whatever coordinates were most useful for describing the system, increasing the simplicity, elegance and scope of the mathematics.

4 Eighteenth and Nineteenth Century Astronomy

4.1 Analysis: Laplace to Clausius

But of course in another obvious sense, the very complexity of the object, the solar system, forced the development of physics, since the solar system was the only thing that could be studied as a celestial mechanical system by the instruments available at the time. The main features of that complexity were already apparent to everyone: around the sun there are many planets, with moons around some planets. Uranus was identified by the important astronomer Herschel in 1781, and the asteroid belt between Mars and Jupiter was correctly identified at the beginning of the 19th century. Moreover, there were no important advances in telescopy until the mid-19th century, so the controversies and advances apropos the mathematical models were notably theoretical and analytic. The culmination of these developments was the publication of Pierre-Simon Laplace's five volume *Mécanique céleste* (1799–1825), where with immense mathematical skill Laplace further elaborated these results into analytical methods for calculating the motions of the planets. In the early 1830s, William Rowan Hamilton discovered that if we regard a certain integral as a function of the initial and final coordinate values, this "principal function" satisfies two first order partial differential equations; Carl Jacobi showed how to use Hamilton's approach to solve dynamical ordinary differential equations in terms of the Hamilton-Jacobi equation, later simplified and generalized by Alfred Clebsch Fraser (2000).

Hermann von Helmholtz's publication of *On the Conservation of Force* (with force (*Kraft*) defined in such a way that we would now translate it as energy) in 1847 was the culmination of efforts to find a mechanical equivalent for heat in the new domain of thermodynamics, and to integrate a theory of mechanics, heat, light, electricity and magnetism by means of the notion of energy, rather than gravitational force. Rudolf Clausius reformulated the work of Sadi Carnot and introduced the second law of thermodynamics in 1850, as well as the notion of entropy in 1865. In an 1870 lecture entitled "On a Mechanical Theorem Applicable to Heat," he introduced the Virial Theorem, which states that in an assemblage of particles in gravitationally bound, stable statistical equilibrium, the average kinetic energy is equal to half the average potential energy. Whereas measuring the potential energy of a system requires the ability to measure its mass, measuring the kinetic energy depends on the measurement of the motions of bodies in the system. In the case of

astronomical bodies, it is much easier to measure the latter than the former, so the Virial Theorem came to assume an important role in twentieth century cosmology, when it was applied to galaxies and galaxy clusters. However, in 1870, these objects were barely discernible: they were referred to as *nebulae*, clouds, because that was how they appeared. Many astronomers supposed that they would prove to have interesting internal structure, after Laplace in 1796, following the speculations of Kant, proposed the nebular hypothesis that the solar system emerged from a cloud of swirling dust. Thus here the issue of models for the heavens reverts to the problem of reference, and the work of the astronomers Sir William Herschel and William Parsons, Earl of Rosse.

4.2 Reference: Messier, Herschel and the Earl of Rosse

The path from the detection of 'nebulae' as cloudy smudges within the sole 'island galaxy' of the Milky Way to the recognition that many of them were in fact other galaxies far distant from our own, with complex internal structure encompassing hundreds of billions of stars is long and winding. Charles Messier catalogued the closest galaxy Andromeda as M31 in 1764, and William Herschel estimated that it was about 2000 times further away from us than Sirius (which is one of the stars closest to us). Herschel's large reflecting telescopes produced a dramatic increase in the ability of astronomers to watch the heavens; in 1789 he proposed that nebulae were made up of self-luminous nebulous material. He made hundreds of drawings of them, looking for significant morphological differences, or patterns of development, as he searched for evidence of his nebular hypothesis that clusters of stars may be formed from nebulae. (Laplace modified the nebular hypothesis, as noted above, to speculate that the solar system was originally formed from a cloud of gases.) Herschel's son John revised his father's catalogue for the Northern hemisphere, and established a catalogue for the Southern Hemisphere as well, and kept alive the question of the composition of the nebulae: what were they made of? Alongside tabulations of positions, astronomical observations were drawn by hand; John Herschel was known for his meticulous sketches, which he hoped could be used in series, and by future astronomers, to determine change and motion in celestial configurations (Nasim 2010).

In 1845, William Parson, Earl of Rosse, built the largest telescope in the world: its speculum mirror was six feet in diameter, with a focal length of over four feet. He hoped to discover some of the fine structure of Herschel's nebulae. Soon after the telescope was set up, next to a smaller one that was equatorially mounted, he pointed it at Messier 51 (what we now call the Whirlpool Galaxy, a bright, face-on spiral with a companion) and discovered both its spiral, swirled structure and its companion. The discernment of the *shape* of the nebula was decisive. He sketched it repeatedly, in two steps: first he used the smaller telescope to scale the drawing, and then the large one to fill in the details. Herschel saw Rosse's sketches, presented at a meeting of the British Association for the Advancement of Science and was

Fig. 4 Rosse's sketch of Messier 51

enthusiastically supportive. These drawings were later improved and engraved, so that the nebula was represented in negative, as black on a white background. So in Rosse's research project, the production of an astronomical image was an interplay between what was seen through the telescope, and what was carefully sketched by lamplight by Rosse and various assistants, thereafter to be re-fashioned as an engraving. This was the first model of a galaxy (Nasim 2010) (Fig. 4).

5 Twentieth Century Astronomy

5.1 From Rosse to Hubble

In the last two decades of the 19th century, astronomers solved various technical problems (for example, how to keep a linked telescope and a camera with a certain required exposure time pointing in the right direction for the required stretch of time) and profited from the introduction of dry plate photography, so that by 1887 a consortium of twenty observatories could produce a comprehensive astronomical atlas from photographic images. Comparison of photographs rapidly made clear how variable sketches had been as records of celestial objects, especially nebulae. Once astronomers had a firmer grasp of what they were trying to look at, the next step was to estimate how far away they were, and then to combine that knowledge with star counts and further estimations of stellar velocities within a given galaxy.

Up to this point, the application of classical mechanics to these mysterious objects had really only been a pipedream. In the first decades of the twentieth century it became a true research program, once nebulae were acknowledged to be extra-galactic objects, much larger and farther away than anyone in the 19th century suspected, and ever more powerful telescopes were able to track their motions and resolve their images. In the meantime, however, classical mechanics was being transformed, and the ensuing theoretical disputes affected the work of astronomers as well.

Thus the development of Newtonian mechanics was *not* "normal science" in Kuhn's sense. The emergence of electro-magnetic theory, the independent development of chemistry, and the study of thermodynamics were shaped by a growing awareness that in different domains forces other than gravity were important and demanded codification, and that the notation of differential equations, the study of symmetries, and the category of energy (as opposed to force) should be central to mechanics. However, the most direct challenge to Newtonian mechanics came from Einstein's special and general theories of relativity, which explored the consequences of the equivalence of inertial frames (special relativity) and of accelerated frames (general relativity), given the constant speed of light. Einstein proposed an equivalence between matter and energy, a 4-dimensional space-time continuum curved locally and perhaps globally by the matter and energy located in it, a dilation or slowing down of the passage of time experienced by items moving close to the speed of light, and the notion of a light-cone as a formal limit on cosmic causal interaction. It was clear that these revisions of classical mechanics would have significant consequences for astronomy, certain aspects of which were beginning to change into modern scientific cosmology. In the late eighteenth and early nineteenth century, cosmology had remained merely speculative, driven by the metaphysical certainty of Leibniz and Goethe that in nature, "everything strives." However, relativity theory did not impinge immediately on the study of galaxies; rather, it was the study of the 'red shift' of the electro-magnetic radiation emitted from stars and the characterization of 'Cepheid variables,' both more closely related to problems of reference and taxonomy than to theoretical speculation, which moved the study of galaxies into the heart of modern cosmology.

The astronomer Edwin Hubble studied galaxies by analyzing the emission spectra of the light emitted by their stars; he noted that the standard patterns of spectral lines were shifted toward the red end of the spectrum. This he interpreted as analogous to a 'Doppler shift,' which we know from ordinary experience as the lowering of the tone (due to the sound wave's perceived lengthening) of a train whistle when the train rushes past us; this effect holds not only for sound waves but also for light waves. That is, Hubble took this 'red shift' as evidence that the galaxies (most of them) were receding from us. His famous law proposed in 1929 posits a linear relation defined by the Hubble constant between recessional velocity and distance, so that a measurement of red shift could be used to give an accurate estimate of how far away from us a galaxy lies. He also used 'standard stars' called Cepheid variables, whose period of variation and absolute luminosity are tightly related, as signposts; in combination, these factors allowed him to see that nebulae

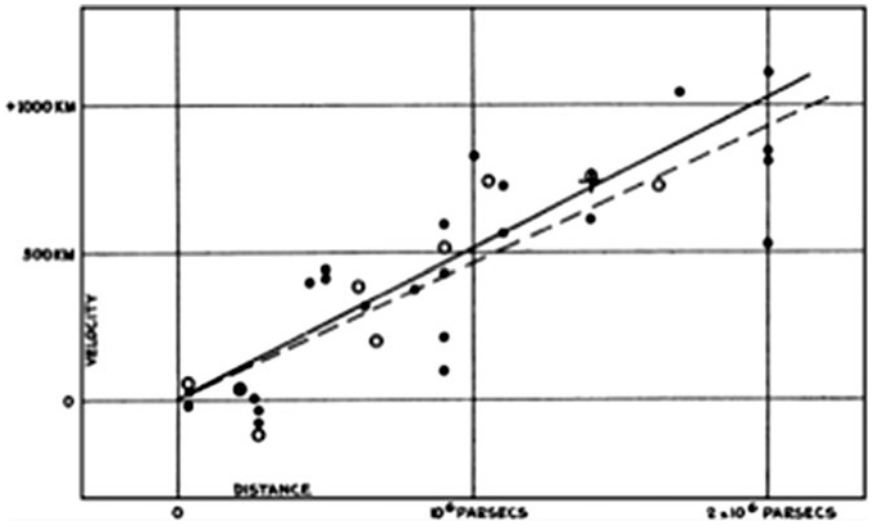


Fig. 5 Hubble's data (1929)

were extra-galactic, and to estimate their distances from us. Thus it was only during the 1920s that the scale of the universe began to dawn on astronomers (Liddle 2009). In 1936, Hubble wrote in his influential book *The Realm of the Nebulae* that "valuable information has been assembled concerning the scale of nebular distances, the general features of nebulae, such as their dimensions, luminosities, and masses, their structure and stellar contents, their large-scale distribution in space, and the curious velocity-distance relation." (Hubble 1982) (Fig. 5).

From that point on, scientists were puzzled about how to address the mismatch between astrophysical theory, originally based on the behavior of objects in the solar system, and the measurement of celestial systems. The orbital speeds of the stars in galaxies should be determined by the total mass of the galaxy pulling on them, and should diminish in proportion to their distance from the center; but stars on the outskirts of galaxies go much too fast. The laws of Newtonian physics predict that such high speeds would pull the galaxy apart. Thus in order to explain the stability of a galaxy, scientists either had to assume there is much more matter in a galaxy than we can see in the form of stars like our sun, or that Newton's laws must be revised for large systems.

5.2 The Dispute Between Zwicky and Hubble

In 1937, the astronomer Fritz Zwicky took issue with Hubble on a number of points. He announced at the beginning of his paper "On the Masses of Nebulae and Clusters of Nebulae," that the determination of the masses of extragalactic nebulae was a

central problem for astrophysics. “Masses of nebulae until recently were estimated either from the luminosities of nebulae or from their internal rotations,” he noted, and then asserted that both these methods of reckoning nebular masses were unreliable. The adding up of observed luminosities gave figures that are clearly too low; and the models used for reckoning mass on the basis of observed internal motions were too indeterminate. Better models were needed, not least because Zwicky was convinced that in addition to luminous matter, galaxies (and the larger formations of galaxy clusters) included ‘dark matter.’ He wrote, “We must know how much dark matter is incorporated in nebulae in the forms of cool and cold stars, macroscopic and microscopic solid bodies, and gases.” (Zwicky 1937). It would be anachronistic to read Zwicky here as supporting or even introducing the current hypothesis of ‘dark matter,’ since he used the term simply to indicate that he thought that our telescopes cannot see some or most of what is actually included in a galaxy or galaxy cluster. There was luminous matter, which we can detect, and dark matter which (as yet) we can’t. This made it all the more important to be able to estimate the mass of a galaxy or galaxy cluster on the basis of the internal movements of its visible components; thus we would have to improve upon the mechanical models used, so that those estimates could become more accurate. He discussed four kinds of models, the first of which, Hubble’s model, he dismissed.

In *The Realm of Nebulae*, Hubble argued that from observations of internal rotations, good values of the mass of a galaxy should be derived. He wrote, “Apart from uncertainties in the dynamical picture, the orbital motion of a point in the equatorial plane of a nebula should be determined by the mass of material inside the orbit. That mass can be calculated in much the same way in which the mass of the sun is found from the orbital motion of the earth (or of the other planets).” (Hubble 1982). However, he expressed some doubts about how to interpret available data about both galaxies and galaxy clusters. Zwicky diagnosed the problem in terms of the indeterminacy of the mechanical model, for one could make the assumption either that the ‘internal viscosity’ of a nebula was negligible, or that it was very great. In the former case, the observed angular velocities will not allow the computation of the mass of the system; in the latter case, the nebula will rotate like a solid body, regardless of what its total mass and distribution of that mass may be. For intermediate and more realistic cases, Zwicky argued, “it is not possible to derive the masses of nebulae from observed rotations without the use of additional information.” If, for example, there were a central, highly viscous core with distant outlying, little-interacting components, one would need information about that viscosity and about the distribution of the outlying bodies. And he dismissed the analogy with the solar system as superficial.

Zwicky went on to propose three other possible models for calculating the mass of a galaxy or galaxy cluster. The second approach was to apply the Virial Theorem. If a galaxy cluster such as the Coma cluster was stationary, then “the virial theorem of classical mechanics gives the total mass of a cluster in terms of the average square of the velocities of the individual nebulae which constitute this cluster.” He argued that the Virial Theorem would work for the system, even if the nebulae are not evenly distributed throughout the cluster. But what if the cluster

was not stationary? A brief calculation showed that, given the velocities, the virial theorem predicts that ultimately it will fly apart, which is odd, since then there should be no galaxy clusters at all; so there must be 'internebular material,' whose nature and density should be further studied. Zwicky concluded that "the virial theorem as applied to clusters of nebulae provides for a test of the validity of the inverse square law of gravitational forces," because the distances are so enormous and these clusters are the largest known aggregates of matter. He also remarked that it would be desirable to apply the virial theorem to individual galaxies, but that it was just too difficult to measure the velocities of individual stars, as it was at that point in time. He treated this practical limitation as if he could not foresee its resolution (Zwicky 1937).

The next model was that of gravitational lensing, a direct application of Einstein's theory of General Relativity; however, this was merely a speculative proposal, and wasn't carried out observationally until 1979. The final model was an extrapolation of ordinary statistical mechanics, "analogous to those which result in Boltzmann's principle." Zwicky's motivation in this section seemed to be to find a theory that would explain large-scale features of the universe without resorting to the kind of cosmological account (like the Big Bang theory, with which Hubble's Law became associated) he opposed, given his general disapproval of Hubble. Zwicky concluded, "It is not necessary as yet to call on evolutionary processes to explain why the representation of nebular types in clusters differs from that in the general field. Here, as in the interpretation of other astronomical phenomena, the idea of evolution may have been called upon prematurely. It cannot be overemphasized in this connection that systematic and irreversible evolutionary changes in the domain of astronomy have thus far in no case been definitely established." For Zwicky, part of what was at stake was whether our model of the whole cosmos should be evolutionary or not (Zwicky 1937).

5.3 Vera Rubin

Thus at the end of the 1930s, two important astronomers who had access to the same observational data on the largest material objects in the universe found themselves associated with two radically opposed views on the direction cosmology should take. Yet they were both equally puzzled by the discrepancy in estimates of the mass of these large objects. The evidence provided by star-counting or galaxy-counting, and the results of mechanically plausible models that calculate mass on the basis of the motions of stars in galaxies and of galaxies within clusters, simply did not agree. So the choice of theory could not be determined by observational results, and the clash of observational results could not be reconciled by theory. A quarter century later, astronomers were finally in a position to measure the velocities of components of a galaxy, and so to calculate the mass of the galaxy. Astronomers already had reliable evidence that a galaxy rotates about its center, based on the gradient in the stellar absorption lines on the major axis and the lack of

such a gradient on the minor axis. If a galaxy were a mechanical system like the solar system, then we should expect that the velocity of its outer regions should decrease, as Kepler (and then, generalizing, Newton and Clausius) demonstrated. The longer periods of revolution of Jupiter and Nepture, and the shorter periods of Mercury and Venus, can be accurately predicted. Even such a distinguished astronomer as Vesto Slipher (1875–1969) continued to characterize the radial velocity data of Andromeda and the Sombrero galaxy as "planetary" into the 1950s.

In the early 1960s, Vera Rubin and her graduate students made careful studies of the velocities of stars on the outskirts of Andromeda, because Rubin was interested in where galaxies actually end; they found that the galaxy rotation curve did not diminish, as expected, but remained flat. In 1970, she and W. Kent Ford, Jr. reported new data on Andromeda, profiting from the identification of almost 700 individual emission regions, as well as the use of image intensifiers that reduced observation times by a factor of 10. The edges of Andromeda did not move slower; they moved just as quickly as the inner regions. (The Galaxy Andromeda is M31 in the Messier Catalogue.) (Rubin and Ford 1970) (Fig. 6).

In 1980, with W. Kent Ford and Norbert Thonnard, she reported similar data for 21 further galaxies. While in the earlier papers she was reticent about drawing explicit conclusions, in this paper she writes, "Most galaxies exhibit rising rotational velocities at the last measured velocity; only for the very largest galaxies are the

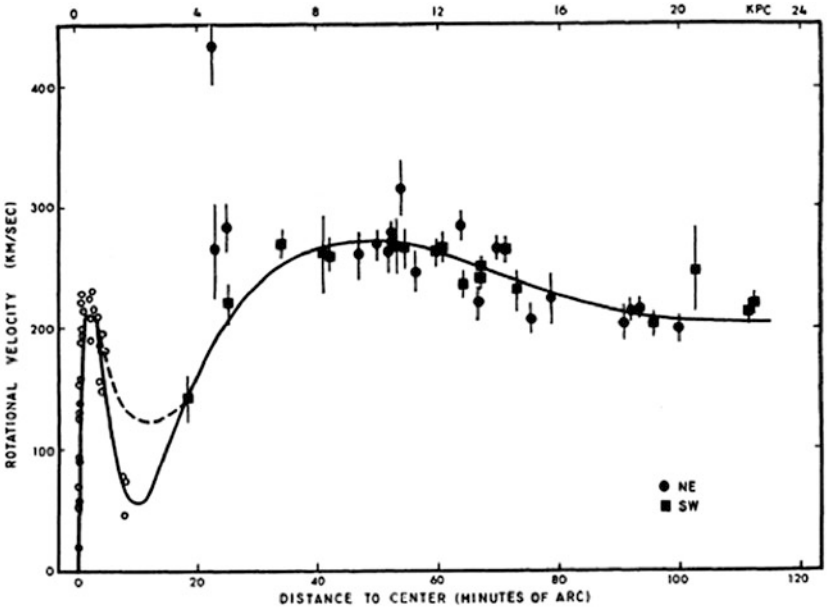


FIG. 9.—Rotational velocities for OB associations in M31, as a function of distance from the center. *Solid curve*, adopted rotation curve based on the velocities shown in Fig. 4. For $R \leq 12'$, curve is fifth-order polynomial; for $R > 12'$, curve is fourth-order polynomial required to remain approximately flat near $R = 120'$. Dashed curve near $R = 10'$ is a second rotation curve with higher inner minimum.

Fig. 6 Rubin's data

rotation curves flat. Thus the smallest Sc's (i.e. lowest luminosity) exhibit the same lack of Keplerian velocity decrease at large R as do the high-luminosity spirals. This form for the rotation curves implies that the mass is not centrally condensed, but that significant mass is located at large R. The integral mass is increasing at least as fast as R. The mass is not converging to a limiting mass at the edge of the optical image. The conclusion is inescapable that non-luminous matter exists beyond the optical galaxy." Since then, her observations have proved consistent with the measurement of velocities in a wide variety of other galaxies (Rubin et al. 1980).

6 Coda

Scientists remain divided about how to address the mismatch between astrophysical theory, originally modeled on our solar system, and data on larger systems. Proponents of the abductive thesis of dark matter argue that, in order to explain the stability of a galaxy or a galaxy cluster, we have to assume that there is much more matter in a galaxy than we can see in the form of stars like our sun. We must posit a spherical halo of dark matter around the spiral or ellipsoid or spheroid that we see. Other scientists are unhappy with a scientific theory based on something that (up till now) has resisted detection altogether. The research program MOND (Modified Newtonian Dynamics) proposes instead that we revise Newtonian Mechanics to explain the uniform velocity of rotation of galaxies. Since its inception thirty years ago, proponents have_ tried various adjustments and refinements, without winning general acceptance. So it seems that we must choose between ad hoc adjustment of principles, or postulating a new kind of matter we can't detect. Clearly, a new kind of model is called for, which will bring reference and theoretical analysis into novel and more fruitful alignment.

References

Diacu, F., Holmes, P.: Celestial Encounters: The Origins of Chaos and Stability, Chaps. 1 and 4. Princeton University Press, Princeton (1996)

Fraser, C.: Hamilton-Jacobi Methods and Weierstrassian field theory in the calculus of variations. In: Grosholz, E., Breger, H. (eds.) The Growth of Mathematical Knowledge, pp. 93–101. Kluwer, Dordrecht (2000)

Grosholz, E.: Representation and Productive Ambiguity in Mathematics and the Sciences, Chap. 7. Oxford University Press, Oxford (2007)

Hendry, R.: Mathematics, representation, and molecular structure. In: Klein, U. (ed.) Tools and Modes of Representation in the Laboratory Sciences, pp. 221–236. Kluwer, Dordrecht (2001)

Hubble, E.: The Realm of the Nebulae, p. 181. Yale University Press, New Haven, Ct. (1982/2013)

Kuhn, T.: The Copernican Revolution, Chap. 6. Harvard University Press, Cambridge, Ma. (1957/2003)

Leech, J.W.: Classical Mechanics, Chap. 3–6. Methuen, London (1958)

Liddle, A., Loveday, J.: Oxford Companion to Cosmology, pp. 165–170. Oxford University Press, Oxford (2009)

Nasim, O.W.: Observation, working images and procedures: the 'Great Spiral' in Lord Rosse's astronomical records and beyond. Br. J. Hist. Sci. **43**(3), 353–389 (2010)

Newton, I.: Mathematical Principles of Natural Philosophy and His System of the World, Motte, A. (trans.), Cajori, F. (ed.) University of California Press, Berkeley (1966)

Rubin, V., Ford, W.K.: Rotation of the Andromeda Nebula from a spectroscopic survey of emission regions. Astrophys. J. **159**, 379–403 (1970)

Rubin, V., Ford, W.K., Thonnard, N.: Rotational properties of 21 Sc galaxies with a large range of luminosities and radii, from NGC 4605 (R = 4 kpc) to UCG 2885 (R – 122 kpc). Astrophys. J. **238**, 471–487 (1980)

Wilson, R.N.: Reflecting Telescope Optics I: Basic Design Theory and Its Historical Development, Chaps. 1.1 and 1.2. Springer, Heidelberg (2014)

Zwicky, F.: On the masses of nebulae and of clusters of nebulae. Astrophys. J. **86**(3), 217–246 (1937)

Models of Science and Models in Science

Carlo Cellucci

1 Premise

With regard to science, one may speak of models in different senses. The two main ones are models of science and models in science. A model of science is a representation of how scientists build their theories, a model in science is a representation of empirical objects, phenomena, or processes of some area of science.

In this article I will describe and compare four models of science: the analytic-synthetic model, the hypothetico-deductive model, the semantic model, and the analytic model. Then I will briefly discuss to what extent each of these models of science is capable of accounting for models in science.

2 The Analytic-Synthetic Model

The analytic-synthetic model of science was introduced by Aristotle and is Aristotle's model of science.

According to Aristotle, "the process of knowledge proceeds from what is more knowable and clearer to us to what is clearer and more knowable by nature" (Aristotle, *Physica*, A 1, 184 a 16–18). Now, what is more knowable and clearer to us is the conclusion we want to establish, while what is clearer and more knowable by nature are the prime premises, or principles. Thus, according to Aristotle, the process of knowledge proceeds from the conclusion we want to establish to the prime premises, or principles.

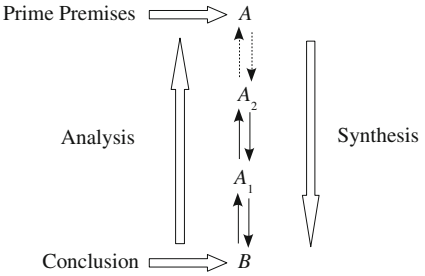
C. Cellucci (✉)
Sapienza University of Rome, Via Carlo Fea 2, Rome, Italy
e-mail: carlo.cellucci@uniroma1.it

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_6

Starting from the conclusion we want to establish, we must find "the necessary premises through which the syllogisms come about" (Aristotle, *Topica*, Θ 1, 155 b 29). We will find them "either by syllogism"—namely by Aristotle's procedure for finding the premises of a syllogism, given the conclusion—"or by induction" (Aristotle, *Topica*, Θ 1, 155 b 35–36). For a detailed description of Aristotle's procedure for finding the premises of a syllogism, given the conclusion, see Cellucci (2013, Chap. 7).

When the premises are found, we "should not put these forward right away, but rather should stand off as far above them as possible" (ibid., Θ 1, 155 b 29–30). Namely, we should find other premises from which the previous premises can be deduced. And so on, until we arrive at premises which are prime premises. The prime premises are principles, because "I call the same thing prime and principle" (Aristotle, *Analytica Posteriora*, A 2, 72 a 6–7). Being principles, the prime premises must be indemonstrable and true. Moreover, they must implicitly contain all about the kind with which they are concerned, and must be of the same kind as the conclusion, because "the indemonstrables," namely the principles, "must be in the same kind as the things demonstrated" (ibid., A 28, 87 b 2–3).

When we arrive at prime premises, the upward process terminates. This is analysis, and the upward process is the analytic method. At this point we try to invert the process, deducing the conclusion we want to establish from the prime premises, thus producing a demonstration of it and hence scientific knowledge. For "to know scientifically is to know through demonstration" (Aristotle, *Analytica Posteriora*, A 2, 71 b 18). This is synthesis, and the deduction process is the synthetic method.



In addition to being true, the prime premises must be known to be true, otherwise we could not say that we "have scientific knowledge of what follows from them, absolutely and properly" (ibid., A 3, 72 b 14). Then the question arises how the prime premises become known to be true. Now, the prime premises cannot become known to be true by demonstration, otherwise they would be demonstrable. But the prime premises, being principles, "are indemonstrable, therefore it will not be scientific knowledge but intuition [*nous*] that is concerned with the principles" (Aristotle, *Magna Moralia*, A 34, 1197 a 22–23). Thus it will be "intuition [*nous*] that apprehends the principles" (Aristotle, *Analytica Posteriora*, B 19, 100 b 12).

Since, for Aristotle, the prime premises become known to be true by intuition, intuition plays an essential role in Aristotle's analytic-synthetic model, being the

way we apprehend the principles. However, the prime premises are not discovered by intuition. They are discovered either by Aristotle's procedure for finding the premises of a syllogism, given the conclusion, or by induction.

3 The Analytic-Synthetic Model and Modern Science

According to a widespread opinion, "modern science owes its origins and present flourishing state to a new scientific method which was fashioned almost entirely by Galileo Galilei" (Kline 1985, p. 284). This opinion, however, is unjustified. The initiators of modern science, from Galileo to Newton, did not fashion a new scientific method, on the contrary, they followed Aristotle's analytic-synthetic model of science.

For example, Newton states that "as in mathematicks, so in natural philosophy, the inquiry of difficult things by the method of analysis, ought ever to precede the method" of synthesis, or "composition" (Newton 1952, p. 404). Now, "analysis consists in making experiments and observations, and in drawing general conclusions from them by induction" (ibid.). More precisely, in analysis "particular propositions are inferred from the phenomena, and afterwards rendered general by induction" (Newton 1962, II, p. 676). Of course, "the arguing from experiments and observations by induction" is "no demonstration of general conclusions; yet it is the best way of arguing which the nature of things admits of" and "may be looked upon as so much the stronger, by how much the induction is more general" (Newton 1952, p. 404).

Once a general conclusion has been reached, "if no exception occur from phenomena, the conclusion may be pronounced generally" (ibid.). On the other hand, if "any exception shall occur from experiments," the conclusion will "then begin to be pronounced with such exceptions as occur" (ibid.). For "in experimental philosophy, propositions gathered from phenomena by induction, when no contrary hypotheses are opposed, must be considered to be true either exactly or very nearly, until other phenomena occur by which" such propositions "are made either more exact or liable to exceptions" (Newton 1962, II, p. 400). Then the propositions will be pronounced with the exceptions.

By this way of inquiry we proceed "from effects to their causes, and from particular causes to more general ones, till the argument end in the most general. This is the method of analysis" (Newton 1952, p. 404). On the other hand, "synthesis consists in assuming the causes discovered, and established as principles, and by them explaining the phenomena proceeding from them, and proving the explanations" (ibid., pp. 404–405).

Thus, according to Newton, premises are obtained from experiments and observations by induction. The latter is not demonstration, and yet is the best way of dealing with the objects of nature, as distinguished from mathematical objects. Moreover, induction is so much stronger by how much it is more general. Premises obtained by means of it are accepted as long as no counterexample occurs. This is

analysis. From the premises, which give the causes, one then deduces the phenomena. This is synthesis.

Clearly, this is Aristotle's analytic-synthetic model of science, except that Newton does not mention Aristotle's procedure for finding the premises of a syllogism, given the conclusion, but only induction.

The method just described is the method by which Newton proceeds. Indeed, he explains that his own propositions "were invented by analysis. But, considering" that ancient mathematicians "admitted nothing into geometry before it was demonstrated by composition," namely by synthesis, Newton composed, that is, he wrote synthetically, what he had "invented by analysis, to make it geometrically authentic and fit for the publick" (Cohen 1971, p. 294). Newton "could have written analytically" what he "had found out analytically," but he "was writing for scientists steeped in the elements of geometry"—namely in Euclid's *Elements*—and was "putting down geometrically demonstrated bases for physical science" (Newton 1967–1981, VIII, p. 451). Therefore, he wrote "in words at length after the manner of the ancients" (Cohen 1971, p. 294). That is, synthetically.

Admittedly, "if any man who understands analysis will reduce the demonstrations of the propositions from their composition back into analysis," he "will see by what method of analysis they were invented" (ibid.). But this will require considerable skill, because synthesis tends to hide analysis. This "makes it now difficult for unskilful men to see the analysis by which those propositions were found out" (ibid., p. 295).

Thus Newton makes it quite clear that he discovered his results by analysis. But he presents them hiding analysis and disclosing only synthesis, in order to make them fit for a public used to the didactic style of Euclid's *Elements*. Thus Newton proceeds in an opposite way with respect to Descartes who, in his *Geometry*, lays down no definitions, postulates and common notions in the beginning, and presents only analysis. Indeed, according to Descartes, in synthesis "there is no difficulty, except in deducing the consequences properly; which" however "can be done even by the less attentive people, provided only that they remember what has gone before" (Descartes 1996, VII, pp. 156–157). Perhaps Newton refers to Descartes when he states that "the mathematicians of the last age have much improved analysis, but stop there and think they have solved a problem when they have only resolved it" by the method of analysis, thus "the method of synthesis is almost laid aside" (Cohen 1971, p. 294).

4 The Disappearance of Analysis

Contrary to what happened in the age of Descartes and Newton, in the period from the eighteenth century to the second half of the nineteenth century, the analytic part of the analytic-synthetic model regresses and ultimately disappears, only the synthetic part of the analytic-synthetic model remains.

This does not mean that, in the period in question, there are no explicit uses of the analytic-synthetic model. An example is provided by Riemann who states that, "for the physiology of a sense organ," in particular the ear, "there are two possible ways of gaining knowledge about its functions. Either we proceed from the constitution of the organ, and from there seek to determine the interaction of its parts as well as its response to external stimuli; or we begin with what the organ accomplishes and then attempt to account for this" (Riemann 1892, p. 338). Now, "by the first way we argue from given causes to effects, by the second way we seek causes of given effects. With Newton and Herbart, we may call the first way synthetic, the second analytic. The first way is closest to the anatomist" (ibid.). On the other hand, "by the second way we seek to give an explanation for what the organ accomplishes. This undertaking can be broken down into three parts. (1) The search for a hypothesis which is sufficient to explain what the organ accomplishes. (2) The investigation of the extent to which this explanation is a necessary one. (3) The comparison with experience in order to confirm or correct such explanation" (ibid., p. 339). The search of a hypothesis is carried out with "the use of analogy" (ibid., p. 341). Both ways are indispensable, because "every synthesis rests upon the results of a preceding analysis, and every analysis, in order to be confirmed or corrected through experience, requires a subsequent synthesis" (ibid., p. 340).

Clearly, this is Aristotle's analytic-synthetic model of science, except that, like Newton, Riemann does not mention Aristotle's procedures for finding the premises of a syllogism given the conclusion, and replaces induction with analogy.

However, despite some exceptions, as already said, in the period from the eighteenth century to the second half of the nineteenth century, the reference to analysis declines and ultimately disappears. At the origin of this there are at least two factors.

One factor is Romanticism, which exalts genius and the role of intuition. Thus Novalis states that scientific discoveries "are leaps—(intuitions, resolutions)" and products "of the genius—of the leaper *par excellence*" (Novalis 2007, p. 28). Therefore the analytic part of the analytic-synthetic model should be abandoned and only the synthetic part should be preserved, because "genius is the synthesizing principle" (ibid., 215). In mathematics "a true method of progressing synthetically is the main thing" and this is the "method of the divinatory genius" (ibid., 100). The synthetic method gives "the regulation of genius" (ibid., 164). Admittedly, "the synthetic method" is "the freezing, wilting, crystallizing, structuring and successive method. The analytic method in contrast, is a warming, dissolving and liquefying method. The former seeks the whole, the latter the parts" (ibid., 175). For "the synthetic course proceeds above all from the constituents (or better, from the elements) to the whole," while "the analytic course from the whole to the elements" (ibid., 194). Neverthless, the true method is the synthetic method, because only the synthetic method permits to build a system in an absolutely free way. Therefore "the true philosopher has a synthetic method" (ibid., 73).

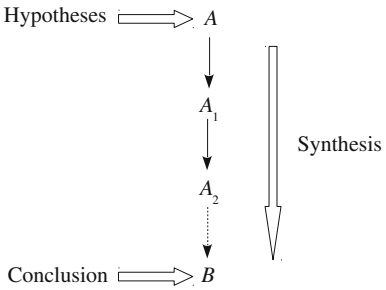
Another factor is the development, in the nineteenth century, of theories involving hypotheses that appeal to unobservable entities and processes, and hence cannot be derived from observation. Thus Whewell states that from these theories it

is clear that "an art of discovery is not possible. At each step of the progress of science, are needed invention, sagacity, genius; elements which no art can give" (Whewell 1847, I, p. viii). Discovery "must ever depend upon some happy thought, of which we cannot trace the origin; some fortunate cast of intellect, rising above all rules. No maxims can be given which inevitably lead to discovery. No precepts will elevate a man of ordinary endowments to the level a man of genius" (ibid., II, pp. 20–21). Since discovery must ever depend upon some fortunate cast of intellect, rising above all rules, Herschel states that "we must not, therefore, be scrupulous as to how we reach to a knowledge of such" theories: "provided only we verify them carefully when once discovered, we must content to seize them wherever they are to be found" (Herschel 1851, p. 164).

Because of these and possibly other factors, at the end of the nineteenth century the analytic part of Aristotle's analytic-synthetic model is abandoned and only the synthetic part is retained. This leads to a new model of science, the hypothetico-deductive model.

5 The Hypothetico-Deductive Model

According to the hypothetico-deductive model of science, formulating a scientific theory about a certain class of phenomena means formulating hypotheses, then deducing consequences from them, and finally comparing consequences with the observation data. The hypotheses are subject to the condition that they must be consistent (namely, non-contradictory) and known to be consistent. The hypotheses are supposed to solve all problems in the relevant field. They can and must be tested by comparing the consequences deduced from them with the observational and experimental data.



Thus Carnap states that formulating a scientific theory about a physical process is "a matter of deducing the concrete sentence which describes the process" from hypotheses, consisting of "valid laws and other concrete sentences. To explain a law," thus a universal fact, "means to deduce it from more general laws" (Carnap 2001, p. 320). There is "great freedom in the introduction" of hypotheses, or "primitive sentences" (ibid., p. 322). However, "every hypothesis must be

compatible with the total system of hypotheses" (ibid., p. 320). That is, it must be consistent with them. Moreover, "the hypotheses can and must be tested by experience" (ibid.). Namely, the consequences deduced from them must agree with the observational and experimental data.

6 The Hypothetico-Deductive Model and Closed Systems

According to the hypothetico-deductive model, a scientific theory is a closed system. This means that the development of the theory remains completely internal to the theory, it involves no interaction with other theories, so a scientific theory is a self-sufficient totality. The hypotheses of the theory are given once for all, and developing the theory only means deducing consequences from them. The consequences contain nothing essentially new with respect to the hypotheses, because deduction is non-ampliative, it simply makes explicit what is implicitly contained in the hypotheses.

Thus, according to the hypothetico-deductive model, a theory is all implicitly contained in its hypotheses. As Kant says, a theory is a whole "all of whose parts still lie very involuted and are hardly recognizable even under microscopic observation" (Kant 1998, A834/B862). It can only "grow internally (*per intus susceptionem*)," that is, by growth from within, and "not externally (*per appositionem*)" (ibid., A833/B861). That a theory is all implicitly contained in its hypotheses is due to the fact that developing a theory means deducing consequences from the hypotheses of the theory, and deduction simply makes "explicit what is implicitly contained in a set of premises. The conclusions to which" deduction leads "assert nothing that is theoretically new in the sense of not being contained in the content of the premises," they can only be "psychologically new" in the sense that we were not aware of their being implicitly contained in the premises, thus we were not aware of "what we committed ourselves to in accepting a certain set of assumptions or assertions" (Hempel 2001, p. 14). As Wittgenstein says, in deduction "there can never be surprises" (Wittgenstein 2002, 6.1251). For in deduction "process and result are equivalent. (Hence the absence of surprise)" (ibid., 6.1261).

7 The Hypothetico-Deductive Model and the Axiomatic Method

Clearly, the hypothetico-deductive model is based on the axiomatic method, more precisely, on what is called the 'modern axiomatic method' to distinguish it from the 'classical axiomatic method'.

Aristotle states the classical axiomatic method by saying that a scientific theory proceeds from hypotheses or axioms "that are true and primitive" (Aristotle, *Analytica Posteriora*, A 2, 71b 20-21). Then it deduces conclusions from them by

"scientific deduction" (ibid., A 2, 71b 18). That is, it deduces conclusions from them by a deduction with true premises. In addition to be true, the hypotheses, or axioms, must be known to be true, otherwise we would not have scientific knowledge of what follows from them. The modern axiomatic method differs from the classical axiomatic method in that it does not require that the hypotheses or axioms be true and known to be true, but only consistent and known to be consistent.

In the classical axiomatic method, that the hypotheses must be true means that they must be true of certain given things and facts. However, it may possible to discover that they are true also of other things and facts. This has led to frame axioms systems in which the hypotheses or axioms are true for a large number of things and facts. A typical example is group theory. Such axiom systems are based on the modern axiomatic method rather than the classical axiomatic method. The difference between the classical axiomatic method and the modern axiomatic method is often viewed as an opposition. Such is the case of the Frege-Hilbert controversy, in which, on the one hand, Frege states: "I call axioms propositions that are true" of certain given things and facts, and "from the truth of the axioms it follows that they do not contradict one another" (Frege 1980, p. 37). Assuming that axioms are true for a large number of things and facts means to detach a scientific theory from reality and "to turn it into a purely logical science" (ibid., p. 43). On the other hand, Hilbert states that "any theory can always be applied to infinitely many systems of basic elements," for example, "all statements of electrostatics hold of course also for any other system of things which is substituted for quantity of electricity," provided "the requisite axioms are satisfied. Thus the circumstance I mentioned is never a defect (but rather a tremendous advantage) of a theory" (ibid., p. 42).

However, as Shoenfield points out, "the difference" between the classical axiomatic method and the modern axiomatic method "is not really in the axiom system, but in the intentions of the framer of the system" (Shoenfield et al. 1967, p. 2).

It must be mentioned that, although Aristotle stated the classical axiomatic method, this is not his model of science, which is the analytic-synthetic model. For him, the classical axiomatic method is only as a model of the teaching of science. Indeed, he states that "didactic arguments are those that deduce from the proper principles of each subject" (Aristotle, *Sophistici Elenchi*, 2, 165 b 1–2). That is, didactic arguments are those based on the classical axiomatic method. As a matter of fact, in his scientific research works Aristotle never uses the classical axiomatic method.

8 Models of Science and Gödel's Incompleteness Theorems

The analytic-synthetic model and the hypothetico-deductive model share a basic limitation: they are incompatible with Gödel's incompleteness theorems.

The analytic-synthetic model is incompatible with Gödel's first incompleteness theorem because, by the latter, for any scientific theory in a given field, satisfying certain minimal conditions, there is a sentence which is true but not deducible from the hypotheses of the theory. This affects the analytic-synthetic model, according to which all true sentences of a theory must be deducible from the hypotheses of the theory. It affects the analytic-synthetic model also because, by Gödel's first incompleteness theorem, the hypotheses from which a given truth is to be deduced need not be of the same kind as that truth, while according to the analytic synthetic model they must be of the same kind.

The hypothetico-deductive model is incompatible with Gödel's first incompleteness theorem because, by the latter, for any scientific theory for a given field satisfying certain minimal conditions, there is a sentence of the theory such that neither A nor $\neg A$ is deducible from the hypotheses of the theory. This affects the hypothetico-deductive model, according to which the hypotheses are supposed to solve all problems in the relevant field.

The analytic-synthetic model and the hypothetico-deductive model are incompatible with Gödel's second incompleteness theorem because, by the latter, for any scientific theory in a given field, satisfying certain minimal conditions, it is impossible to prove, by any reliable means, that the hypotheses of the theory are consistent, even more that they are true. This affects the analytic-synthetic model, according to which the hypotheses of a theory must be true and known to be true. It also affects the hypothetico-deductive model, according to which the hypotheses of a theory must be consistent and known to be consistent.

9 Curry's Alleged Way Out

According to Curry, the analytic-synthetic model and the hypothetico-deductive model may be retained if we assume that scientific knowledge in a given field is not represented by a single theory but rather by a growing sequence of theories. Gödel's first incompleteness theorem only entails that "the concept of intuitively valid proof cannot be exhausted by any single formalization" (Curry 1977, p. 15). But scientific knowledge for a given field can be represented by a growing sequence of theories in which each theory properly includes the preceding one, in the sense that the hypothesis of a theory properly include the hypotheses of the preceding one.

Curry's way out, however, is not viable because, in Curry's growing sequence of theories, proof is not a fixed thing but a growing thing. Indeed, as Curry acknowledges, "proof is precisely that sort of growing thing which the intuitionists have

postulated for certain infinite sets" (ibid.). But this notion of proof is incompatible with the analytic-synthetic model and the hypothetic-deductive model, according to which proof is a fixed thing. Each theory in Curry's sequence is a step in a growing sequence of theories, and, as Gödel points out, "there cannot exist any formalism which would embrace all these steps" (Gödel 1986–2002, II, p. 151). But the existence of such a formalism would be necessary if proof is to be a fixed thing.

10 Other Limitations of the Hypothetico-Deductive Model

In addition to being incompatible with Gödel's incompleteness theorems, the hypothetico-deductive model has other limitations. It leaves to one side the crucial issue of how to find hypotheses, limiting itself to saying that, in order to find them, "creative ingenuity is required" (Carnap 1966, p. 33). But this is a non-explanation, it completely evades the issue. Moreover, it may occur that the observational and experimental data may confirm not only our hypotheses, but also other hypotheses which are incompatible with our hypotheses. The hypothetico-deductive model has no argument to claim that the test confirms our hypotheses in preference to the other hypotheses. Furthermore, the hypothetico-deductive model is incapable of accounting for the process of theory change, that is, the process in which one theory comes to be replaced by another. For according to it, a theory has no rational connection with the preceding one, except that it agrees with more observational and experimental data than the preceding one. Thus the hypothetico-deductive model leaves to one side not only the crucial issue of the discovery of hypotheses, but also the equally crucial issue of the process of theory change.

11 The Semantic Model

In the second half of the twentieth century, the support for the hypothetico-deductive model declines, and this model is gradually replaced by the semantic model. There are some slightly different versions of the semantic model. I will consider van Fraassen's version.

According to the semantic model, to formulate a scientific theory about certain phenomena is to specify a family of models. The concept of model is supposed to be the same in mathematics and the empirical sciences. A model is a structure, consisting of a set along with a collection of operations and relations that are defined on it. A scientific theory is adequate if it has some model which is isomorphic to the phenomena that the theory is intended to theorize.

Thus van Fraassen states that "to present a theory is to specify a family of structures, its models" (van Fraassen 1980, p. 64). Such family of structures is specified "directly, without paying any attention to questions of axiomatizability, in any special language" (van Fraassen 1989, p. 222). Then, if a theory as such is to be

identified with anything at all, it "should be identified with its class of models" (ibid.). Here "a model is a mathematical structure" (van Fraassen 2008, p. 376, Footnote 18). More precisely, "a model is a structure plus a function that interprets the sentences in that structure" (van Fraassen 1985, p. 301). If "a theory is advocated then the claim made is that these models can be used to represent the phenomena, and to represent them accurately," where we say that "a model can (be used to) represent a given phenomenon accurately only if it has a substructure isomorphic to that phenomenon" (van Fraassen 2008, p. 309).

The semantic model, however, has some serious shortcomings. A model is a structure and hence a mathematical object, while the phenomenon is not a mathematical object. Indeed, van Fraassen himself asks: "If the target," that is, the phenomenon, "is not a mathematical object, then we do not have a well-defined range for the function, so how can we speak of an embedding or isomorphism or homomorphism or whatever between that target and some mathematical object?" (ibid., p. 241). His answer is that we compare the model not with the phenomenon but rather with the data model, that is, our representation of the phenomenon. The data model "is itself a mathematical structure. So there is indeed a 'matching' of structures involved" and "is a 'matching' of two mathematical structures, namely the theoretical model and the data model" (ibid., p. 252). But van Fraassen's answer is unsatisfactory, because the data model is a mathematical object while the phenomenon is not a mathematical object, which raises the question of the matching of the data model and the phenomenon. Thus van Fraassen's answer just pushes the problem back one step.

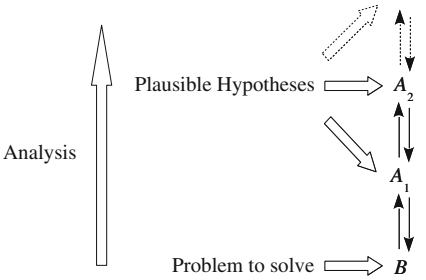
Moreover, even a fiction may have a model, in the sense of a structure. Therefore, it is not models that can make a distinction between fictions and reality.

Furthermore, the semantic model entails that scientific theories, being families of structures, are static things. But scientific theories undergo development. The semantic model has no alternative than treating theory development as a progression of successive families of models. But then the question arises how the transition from a theory to the next one in the progression comes about. The semantic model has nothing to say about this, because it does not account for the process of theory formation, which is essential to explain the development of theories and the process of theory change. Therefore, the semantic model cannot account for the dynamic character of scientific theories. This is a structural limitation of the semantic model.

12 The Analytic Model

An alternative to the above models is the analytic model. According to it, in order to solve a given problem, we start from the problem and look for some hypothesis capable of solving it. The hypothesis is obtained from the problem, and possibly other data already available, by some non-deductive rule—induction, analogy, metaphor, etc. The hypothesis need not belong to the same field as the problem and

must be plausible, that is, in accord with the present experience. But the hypothesis is in its turn a problem that must be solved, and is solved in the same way. That is, we look for another hypothesis from which a solution to the problem posed by the previous hypothesis can be deduced, it is obtained from the latter problem, and possibly other data already available, by some non-deductive rule, it need not belong to the same field as the problem, and must be plausible. And so on, ad infinitum.



Thus, unlike the analytic-synthetic model, the analytic model assumes that analysis is not a finite process terminating with principles, but rather a potentially infinite process which leads to more general hypotheses, thus it is an unending quest.

Moreover, unlike the hypothetico-deductive model, the analytic model establishes a rational connection between subsequent theories. The hypotheses of the new theory are rationally connected with those of the preceding theory because they are formulated through an analysis of the reasons why the hypotheses of the preceding theory are no longer plausible.

13 The Analytic Model and Open Systems

According to the analytic model, a scientific theory is an open system. This means that the development of the theory need not remain completely internal to the theory, it may involve interactions with other theories, so a scientific theory is not a self-sufficient totality.

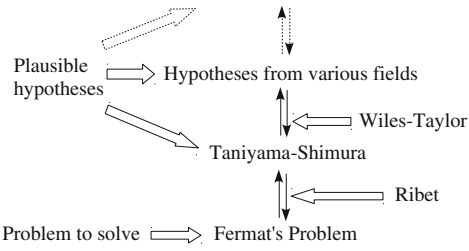
No system of hypotheses may solve all the problems of a given field, any such system is inherently incomplete and must appeal to other systems to bridge its gaps. Therefore the hypotheses of the theory are not given once for all, and developing the theory need not merely mean deducing consequences from them. It may involve replacing the hypotheses with more general ones, obtained through interactions with other scientific theories, according to a potentially infinite process.

14 The Neglect of the Analytic Model

In the last century the analytic model has been generally neglected. One of its few supporters is Pólya, according to whom scientific hypotheses are found "by plausible reasoning" (Pólya 1954, I, p. v). Contrary to deductive reasoning, which is "safe, beyond controversy, and final," plausible reasoning is "hazardous, controversial, and provisional" (ibid.). However, deductive reasoning is "incapable of yielding essentially new knowledge about the world around us. Anything new that we learn about the world involves plausible reasoning" (ibid.). The latter is "the kind of reasoning on which" the "creative work will depend" since to discover hypotheses one has "to combine observations and follow analogies" (ibid., I, p. vi). Thus Pólya shares the basic idea of the analytic model that scientific hypotheses are obtained by logical procedures such as induction, analogy, metaphor, etc.

However, Pólya limits the scope of the analytic model, because he reduces plausibility to probability. Indeed, he states that "the calculus of plausibilities obeys the same rules as the calculus of probabilities" (Pólya 1941, p. 457). This claim is unjustified, because there are hypotheses which are plausible but, in terms of the classical concept of probability, have zero probability. On the other hand, there are hypotheses which are not plausible but, again in terms of the classical concept of probability, have a non-zero probability. The same holds on other concepts of probability, see Cellucci (2013, Sect. 20.4).

That in the last century the analytic model has been generally neglected does not mean, however, that it has not been tacitly or unconsciously used. An example is the solution of Fermat's problem: Show that there are no positive integers x, y, z such that $x^n + y^n = z^n$ for $n > 2$. The problem was solved by Ribet using the Taniyama-Shimura hypothesis: Every elliptic curve over the rational numbers is modular. Indeed Ribet showed: "Conjecture of Taniyama-Shimura ⇒ Fermat's Last Theorem" (Ribet 1990, p. 127). But the Taniyama-Shimura hypothesis was in its turn a problem that had to be solved. It was solved by Wiles and Taylor using plausible hypotheses from various mathematics fields. And so on.



15 The Analytic Model and Gödel's Incompleteness Theorems

While the analytic-synthetic model and the hypothetico-deductive model are incompatible with Gödel's incompleteness theorems, the analytic model is compatible with the latter and even supported by them.

For according to the analytic model, no system of hypotheses can solve all the problems of a given field. The hypotheses are bound to be replaced sooner or later with other more general ones through a potentially infinite process, since every system of hypotheses is incomplete and needs to appeal to other systems to bridge its gaps. Thus the analytic method is supported by Gödel's first incompleteness theorem.

Moreover, according to the analytic method, the hypotheses for the solution to a problem are not definitive, true and certain but only provisional, plausible and uncertain. Thus the analytic method is supported by Gödel's second incompleteness theorem.

16 Models in Science

After considering models of science, I will briefly consider models in science. As already mentioned, a model in science is a representation of empirical objects, phenomena, or processes.

It is out of question that there is an optimal model in science. In the seventeenth century France, the minister Colbert charged the astronomer Gian Domenico Cassini to make an extremely detailed map of France. The map was the work of four different generations of the Cassini family and is so detailed that, as Calvino says, "every forest in France is drawn tree by tree, every church has its bell-tower, every village is drawn roof by roof, so that one has the dizzying feeling that beneath one's eyes are all the trees and all bell-towers and all the roofs of the Kingdom of France" (Calvino 2014, p. 23).

Perhaps this inspired Borges' story about an empire in which "the craft of cartography attained such perfection" that "the college of cartographers evolved a map of the empire that was of the same scale as the empire and that coincided with it point for point" (Borges 1972, p. 141). A previous Lewis Carroll's story goes even further. One of the characters of the story says that in his Kingdom they had "the grandest idea of all" about mapmaking, that is, to make "a map of the country, on the scale of a mile to the mile" (Carroll 1996, p. 556). But "the farmers objected: they said it would cover the whole country, and shut out the sunlight! So we now use the country itself, as its own map, and I assure you it does nearly as well" (ibid., p. 557).

Carroll's story indicates that there cannot be any optimal model in science, because the optimal model of reality would be reality itself. But, contrary to what

the story's character says, using reality itself as its own model would not do nearly as well. As Boltzmann states, "no theory can be objective, actually coinciding with nature," but "each theory is only a mental picture of phenomena" (Boltzmann 1974, pp. 90–91).

In science there cannot be an optimal model, but only models suited to particular purposes. There are several kinds of models suited to particular purposes, such as physical models, scale models, analogical models, mathematical models, just to name a few. It would be impossible to discuss all kinds of models in science here. Instead, I will argue that not all models of science are equally capable of accounting for models in science.

17 The Hypothetico-Deductive Model and Models in Science

The hypothetico-deductive model of science is incapable of accounting for models in science. For according to it, the scientific activity can be exclusively described in terms of deduction of consequences from hypotheses, subject only to the requirement of consistency. Therefore, use of models in science is inessential and can be eliminated.

Thus Carnap states that "it is important to realize that the discovery of a model has no more than an esthetic or didactic or at best a heuristic value, it is not at all essential for a successful application of the physical theory" (Carnap 1939, p. 68).

This contrasts with the fact that, as Morrison and Morgan state, "models are one of the critical instruments of modern science. We know that models function in a variety of different ways within the sciences to help us to learn not only about theories but also about the world" (Morrison and Morgan 1999, p. 10). In particular, models are instruments of discovery.

Moreover, in the hypothetico-deductive model one must always be able to consider arbitrary models of hypotheses, or axioms. Thus, in the case of the axioms of geometry, Hilbert famously stated that, "instead of 'points, straight lines, and planes', one must always be able to say, 'tables, chairs, and beer mugs'" (Hilbert 1970, III, p. 403). This contrasts with the fact that, in the actual practice of science, one does not consider arbitrary models of axioms but only specific ones. In particular, the model in terms of tables, chairs, and beer mugs is never considered. In practice, there are always reasons for considering a model of the axioms rather than another one, and these reasons do not depend on the deductive model but are external to it. Thus the decision to consider a model of the axioms rather than another one cannot be justified in terms of the hypothetico-deductive model.

18 The Semantic Model and Models in Science

One would have thought that, unlike the hypothetico-deductive model, the semantic model would be capable of accounting for models in science. For according to it, formulating a scientific theory about certain phenomena means specifying a family of models.

But it is not so because, as already pointed out, the semantic model is unable to account for the relation between a model and the phenomena and for theory change. The semantic model puts emphasis on the static aspect of physical systems, namely their structure, but physical systems have both structural and behavioral properties. Behavior refers to state transitions and dynamic properties—operations and their relationships. Models should be able to express how and when changes occur to entities and relate with one another. Structures are unable to express that.

19 The Analytic Model and Models in Science

The analytic model is capable of accounting for models in science. According to it, solving a problem involves formulating hypotheses. Now, while many hypotheses in science are expressed using sentences in language, many other hypotheses are expressed using models. Thus models are not ancillary to doing science, but central to the solution of scientific problems. A model is the hypothesis that certain properties of the world can be represented in a certain way for certain purposes.

While, according to the semantic model of science, models are structures, according to the analytic model of science they can be a wide range of things, including words, equations, diagrams, graphs, photographs, computer-generated images, dynamic entities, etc. The question of isomorphism does not arise, because a model is only the hypothesis that certain properties of the world can be represented in a certain way for certain purposes. Thus the analytic model is capable of accounting for models in science.

20 Conclusion

That, unlike the hypothetico-deductive and the semantic model, the analytic model is capable of accounting for models in science, justifies the claim that not all models of science are equally capable of accounting for models in science. Science is a more complex process than the hypothetico-deductive or the semantic model suggest. To account for science it is necessary to account for theory formation and theory change.

Calvino states that “knowledge always proceeds via models, analogies, symbolic images, which help us to understand up to a certain point; then they are discarded,

so we turn to other models," other analogies, other symbolic "images" (Calvino 2014, p. 119). Only the analytic model seems capable of accounting for this dynamic character of scientific knowledge.

References

Boltzmann, L.E.: Theoretical Physics and Philosophical Problems. Selected Writings. Reidel, Dordrecht (1974)

Borges, J.L.: A Universal History of Infamy. Dutton, New York (1972)

Calvino, I.: Collection of Sand. Houghton Mifflin Harcourt, New York (2014)

Carnap, R.: Foundations of Logic and Mathematics. Chicago University Press, Chicago (1939)

Carnap, R.: Philosophical Foundations of Physics. An Introduction to the Philosophy of Science. Basic Books, New York (1966)

Carnap, R.: Logical Syntax of Language. Routledge, London (2001)

Carroll, L.: Sylvie and Bruno Concluded. In: The Complete Illustrated Lewis Carroll, pp. 457–674. Wordsworth, Ware (1996)

Cellucci, C.: Rethinking Logic: Logic in Relation to Mathematics, Evolution, and Method. Springer, Dordrecht (2013)

Cohen, I.B.: Introduction to Newton's 'Principia'. Cambridge University Press, Cambridge (1971)

Curry, H.B.: Foundations of Mathematical Logic. Dover, New York (1977)

Descartes, R.: Œuvres. Vrin, Paris (1996)

Frege, G.: Philosophical and Mathematical Correspondence. Blackwell, Oxford (1980)

Gödel, K.: Collected Works. Oxford: Oxford University Press (1986–2002)

Hempel, C.G.: The Philosophy of Carl G. Hempel. Oxford University Press, Oxford (2001)

Herschel, J.F.W.: Preliminary Discourse on the Study of Natural Philosophy. Longmans, London (1851)

Hilbert, D.: Gesammelte Abhandlungen. Springer, Berlin (1970)

Kant, I.: Critique of Pure Reason. Cambridge University Press, Cambridge (1998)

Kline, M.: Mathematics for the Nonmathematician. Dover, New York (1985)

Morrison, M., Morgan, M.S. (eds.): Models as Mediators. Perspectives on Natural and Social Science. Cambridge University Press, Cambridge (1999)

Newton, I.: Opticks, or a Treatise of the Reflections, Refractions, Inflections and Colours of Light. Dover, New York (1952)

Newton, I.: Mathematical Principles of Natural Philosophy and The System of the World. University of California Press, Berkeley (1962)

Newton, I.: The Mathematical Papers. Cambridge University Press, Cambridge (1967–1981)

Novalis: Notes for a Romantic Encyclopedia. Das Allgemeine Brouillon. State University of New York Press, Albany (2007)

Pólya, G.: Heuristic Reasoning and the Theory of Probability. Am. Math. Mon. **48**, 450–465 (1941)

Pólya, G.: Mathematics and Plausible Reasoning. Princeton University Press, Princeton (1954)

Ribet, K.A.: From the Taniyama-Shimura Conjecture to Fermat's Last Theorem. Annales de la Faculté des Sciences de Toulouse-Mathématiques **11**, 116–139 (1990)

Riemann, B.: Gesammelte mathematische Werke und wissenschaftlicher Nachlass. Teubner, Leipzig (1892)

Shoenfield, J.R.: Mathematical Logic. Addison-Wesley, Reading (1967)

van Fraassen, B.C.: The Scientific Image. Oxford University Press, Oxford (1980)

van Fraassen, B.C.: Empiricism in the Philosophy of Science. In: Churchland, P.M., Hooker C.A. (eds.) Images of Science: Essays on Realism and Empiricism, pp. 245–308. The University of Chicago Press, Chicago (1985)

van Fraassen, B.C.: Laws and Symmetry. Oxford University Press, Oxford (1989)
van Fraassen, B.C.: Scientific Representation: Paradoxes of Perspective. Oxford University Press, Oxford (2008)
Whewell, W.: The Philosophy of the Inductive Sciences, Founded upon Their History. Parker, London (1847)
Wittgenstein, L.: Tractatus Logico-philosophicus. Routledge, London (2002)

Mechanistic Models and Modeling Disorders

Raffaella Campaner

Abstract Recent debate has focused on how disorders should be modeled, and on how their onset, course and final outcome should be explained. I shall here address some issues arising from modeling neuropsychiatric disorders, which are in many cases still poorly understood, subject to a very high rate of individual variations, and tackled from different disciplinary standpoints. After recalling a few core features of current views on mechanistic models, and related views on psychiatric disorders, I shall discuss some models of Attention Deficit Hyperactivity Disorder. The main aspects of such models are analyzed in the light of the philosophical debate on the elaboration and use of mechanistic models, stressing the distance between the two. The paper highlights the many aspects entering the dynamics of modeling disorders and discusses a few problematic issues of explanatory models elaborated in an actual medical scenario that neo-mechanist accounts can only partly capture.

Keywords Mechanisms · Mechanistic models · Explanation

1 Mechanistic Models and Neuropsychiatric Disorders

Given the almost ubiquitous use of mechanistic notions in the biomedical and health sciences, an understanding of the scientific endeavor in these fields requires, amongst others, an understanding of how mechanistic models are conceived, elaborated and employed. While being much indebted to Wesley Salmon's conception of probabilistic mechanism (Salmon 1984, 1998), the last few decades have seen the development of so-labeled "neo-mechanist views" with distinctive features. Different definitions of mechanisms have been put forward by authors like Glennan, Machamer, Darden, Craver, Bechtel, Richardson, Abrahamsen, Tabery, and others.

R. Campaner (✉)
Department of Philosophy and Communication, University of Bologna, Bologna, Italy
e-mail: raffaella.campaner@unibo.it

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_7

Without entering into controversies on the different specific characterizations of mechanisms provided by the various theories, we can adopt a minimal rough notion of mechanism as an organized set of component parts, performing some activities and interacting with each other in the production of some outcome behavior. A mechanism is taken to underlie a given behavior, and is hence to be identified according to the description of the behavior under enquiry. Component parts of the mechanism and their activities, their spatial and temporal organization (position, order, duration, …), and their mutual interactions are held to bring about a given output behavior. A mechanistic explanation outlines the mechanism responsible for the production of the behavior under investigation, by indicating not just its inputs and outputs, but also what occurs *in between* what are regarded the initial causal factors and the final outcome to be explained, thus opening the "black box" of a system's functioning.

Mechanisms are organized systems of interacting parts; *mechanistic models* are accounts of mechanisms. In Glennan's words, "a mechanical model consists of (i) a description of the mechanism's behavior (the behavioral description); and (ii) a description of the mechanism that accounts for that behavior (the mechanical description)" (Glennan 2005, p. 446). Whereas the *behavioral* description is a description of the overall behavior a mechanism brings about, namely of *what the mechanism does*, the *mechanical* description is a description of *how the mechanism produces it*, by the arrangement and working of its parts. The behavioral description hence amounts to the explanandum, and the mechanical description to the explanans: a mere description of the behaviour, without an account of the underlying mechanism, has no explanatory import.

When searching for mechanisms, what we usually obtain are representations with some degree of abstraction—mechanisms' "sketches" and "schemata" in the words of Machamer et al. (2000). A "sketch" is an abstraction for which bottom out entities and activities cannot (yet) be supplied or which contains gaps in its stages. A "mechanism schema" is an abstract description of a type of mechanism that can be filled with already known component parts and their activities. "Mechanism schemata (or sometimes mechanistic models)" are required to be "complete enough *for the purposes at hand*" (Craver and Darden 2013, p. 30, italics added). This in turn seems to require that the purpose of the enquiry be already clear and explicit when the mechanism is sought and the schema put forward, which might not always be the case. Craver (2006) has suggested further concepts to differentiate between mechanistic models, which attempt to provide a more adequate rendering of the dynamic modeling process. *How-possibly models* are only loosely constrained conjectures on the mechanism that produces the phenomenon at stake, with both the existence of the conjectured parts and their engagement in the conjectured activities being highly uncertain.[1] Instead, h*ow-actually models* describe the

[1]For instance, "some computer models are purely how-possibly models" (Craver 2006, p. 366).

components, activities, and organizational features of the mechanism that are *as a matter of fact* involved in the production of the phenomenon: they illustrate how a mechanism actually works, not just merely how it could work if the posited entities and activities existed. How-possibly models explain how some output could be produced; they provide some explanatory, yet just conjectural information on the mechanism's functioning. If the mechanism is then discovered to work as described by the how-possibly model, this turns into a how-actually one. *How-plausibly* models lie somewhere in between, and are more or less consistent with the known constraints on features of components, activities and organization, triggering and inhibiting conditions of the target system.

Some of these conceptual tools have recently been deemed useful to deal with medical issues, and, more specifically, mental disorders and some of the epistemological problems they raise. That the mechanistic approach should be seen not as the only, but as one of the appropriate approaches for psychiatry is maintained by an eminent psychiatrist, Kendler (2008a, b),[2] who believes it is naturally suited to a multicausal framework. According to Kendler, mechanistic modeling fits psychiatry insofar as it allows complicated mechanisms to be decomposed into simple subunits, to study them in isolation and then reassemble constituent parts into their functioning wholes. While this operation can be straightforward when dealing with additive mechanisms, it is much more problematic in a field like psychiatry, where the causal networks investigated present multiple nonlinear interactions between biological, psychological and socioeconomic processes, and often causal loops. A significant example can be given by alcohol dependence, whose causal factors include both molecular and genetic factors (e.g. aggregate genetic factors) and social, cultural, and economic components (e.g. drug availability, forms of ethanol commonly consumed in a social group, acceptability of public drunkenness, religious beliefs, level of taxation on alcoholic beverages, sizes of alcoholic beverages containers allowed,…). What is at issue in the construction of a causal picture of mental disorders is that psychiatry does not demand a clarification of biological, psychological or socio-cultural processes per se, but complex systems resulting from some peculiar intertwining of such different kinds of processes, which can impact on each other in various ways. For instance, the actions of biological factors can be modified by the environment (e.g. light-dark cycle), stressful life experiences (e.g. maternal separation), and cultural forces (e.g. the social acceptability or not of a given behaviour). If construed without privileging any single level a priori, without any specific ontological commitment on some single level deemed the most fundamental, and with just a focus on genuinely productive relations, a mechanistic account can provide a middle ground—Kendler suggests—between hard reduction and hard emergence, both to be avoided. Decomposition is claimed to be driven by

[2]Kendler is thinking of a mechanistic approach like William Bechtel's. Kendler also supports, with different motivations, the interventionist view, and suggests an "integrative pluralism" as the most adequate explanatory framework for psychiatry. See Kendler (2005, 2012), Campaner (2014, forthcoming).

a reductionist stance, while theoretical rearrangement of constituent parts and their activities into complex wholes is guided by some sense of high level organization.[3]

How to model the interactions of many diverse factors and reach a correct and adequate causal account is also extremely relevant for clinical purposes. Although mechanistic knowledge is not necessary to implement therapies and preventive policies, it significantly increases successful interventions on both individual cases and at a population level. Amongst others, Murphy (2010, 2011) stresses that in psychiatry—as in other medical fields—we use *models* to explain *exemplars*, which are idealized representations of the symptoms of disorders and their course. Exemplars take collections of symptoms—which can be many and diverse—to unfold over time in analogous ways, and take patients to respond similarly to the same treatments. To explain, we unravel the pathogenic processes accounting for the phenomenon described in the exemplar. In doing so, we appeal to mechanistic knowledge concerning what are regarded as standard forms of behavior of the systems involved—e.g., the standard neurobiological functioning of the brain. Clinical reasoning then "adjusts" exemplars to the real cases it happens to confront, which always differ to some extent from what is assumed as the prototypical representation of the disease. Clinical practice is hence strongly affected by what are taken to be the *standard* forms of behavior of the systems involved, and by knowledge on what are regarded as the underlying mechanisms.

Mechanistic models have also been deemed useful to *define* and *classify* mental disorders. For instance, a recent joint work by Kendler, Zachar and Craver suggests the mechanistic approach can be employed as a tool to identify mental disorders through different social and cultural contexts by focusing on some *shared physiological mechanism*. Disorders ought thus to be defined in terms of mutually reinforcing networks of causal mechanisms. It is acknowledged that explanatory structures underlying most psychiatric disorders are still quite far from being understood and are likely to be messy, and that cultural and social factors significantly shape the disorder concepts. However, the identification of common mechanisms underlying distinct cases is taken as the possible ground for a taxonomy, to cross cultural and historical contexts. According to Kendler, Zachar and Craver, what is needed for classificatory purposes, is "a scientific model […] that accommodates variability in members of the kinds, multiple etiologies, and probabilistic interactions between causes and outcomes" (Kendler et al. 2011, p. 1143). These sound like rather demanding requirements on classificatory practices. While classifications provided by the various versions of the DSM are mainly symptomatic, these authors suggest that some invariant causal relations underlying clusters of symptoms are to be sought which hold over and above historical, cultural and socio-economic contexts and the corresponding classifications.[4] What are taken as clusters of symptoms in different contexts can be produced by the same

[3]We cannot dwell here on reductionist and antireductionist stances in psychiatry. See e.g. Schaffner (2013).

[4]On mechanisms and psychiatric classification, see also Sirgiovanni (2009).

underlying mechanisms, and the same cluster of symptoms can be brought about by different underlying mechanisms in different cases. Mechanisms are sought which work "at multiple levels, including the symptoms themselves, in addition to mechanisms investigated by the molecular, physiological, computational, psychological and social sciences" (Kendler et al. 2011, p. 1148).

2 Neuropsychiatric Models: Examples from ADHD Investigations

Without claiming that a mechanistic approach can provide the only or best account of mental disorders, I shall suggest some reflections on mechanistic models *in scientific practice* by specifically considering the modeling of Attention Deficit Hyperactivity Disorder (ADHD), a disorder which has been the object of increasing attention and is still far from being thoroughly understood. DSM V (2013) defines ADHD as: "a persistent pattern of inattention and/or hyperactivity-impulsivity that interferes with development, has symptoms presenting in two or more settings (e.g. at home, school, or work), and negatively impacts directly on social, academic, or occupational functioning". The very definition—and hence diagnosis—of ADHD has varied significantly over the last decades. Whereas DSM I (1952) made no mention of the disorder, DSM II (1968) introduced an analogous pathology characterized by short attention span, hyperactivity and restlessness, which was labeled "hyperactivity reaction to childhood". The disorder was most strictly linked to hyperactive behavior, and confined to childhood. DSM III (1980) relabeled it ADD, i.e. attention deficit disorder, being regarded as primarily a problem of inattention, rather than hyperactivity.[5] More recently, DSM IV (1994) included ADHD and distinguished three possible presentations: predominantly hyperactive-impulsive, predominantly inattentive, and combined presentation. Symptoms can change over time, and so can the presentation. To be diagnosed with ADHD, symptoms must be present before the age of 12 years. According to DSM V, children must present at least six symptoms from either—or both—the inattention group of criteria and the hyperactivity and impulsivity criteria, while adolescent and adults must present at least five. ADHD has long been regarded as just a childhood psychiatric condition, but it is now recognized to persist into adolescence and adulthood in a large number of cases.

As it appears, definitions rely on symptoms. Symptomatic behavior includes failure to pay close attention to details, difficulties in organizing tasks and activities, excessive talking, deficits in working memory, regulation of motivation and motor control (e.g. inability to remain seated in appropriate circumstances), difficulties in getting started or sustaining efforts for tasks, in modulating experience and

[5]Two subtypes were identified: ADD/H, i.e. with hyperactivity, and ADD/WO, i.e. without hyperactivity.

expressing emotions, in regulating sleep and alertness. Research has largely focused on underlying cognitive difficulties and executive function impairments, regarded as the core features of the pathology. Specific gene variants have been associated with ADHD,[6] and the pathology is currently investigated from different disciplinary standpoints (e.g. neuropsychiatry, psychiatric genetics, psychiatric epidemiology, clinical psychology,…). The etiology is still uncertain, relationships with the environment fairly opaque, and the disorder presents arrays of clinical symptoms, which are treated both pharmacologically and by means of behavioral interventions. The effectiveness of separate and joint pharmacological and behavioral treatment is widely discussed.[7] Among the theoretical models of the disorder that have been elaborated I shall here consider some influential ones grounded on neurobiology. More specifically, I will consider two causal models which account for the disorder in terms of a single core disorder, i.e. *the executive dysfunction model* and *the motivational model*, and some of their developments, conveyed both verbally and visually. They are not the only available models, but have been very successful and have generated lively debate. A close look at their core characteristics can shed light on the features and roles of models in scientific practice, in dealing with disorders whose representation and explanation are still controversial.

The *executive dysfunction model* stresses the role of executive dysfunction due to deficient inhibitory control, which is ascribed to disturbances in the frontodorsal striatal circuit and associated mesocortical dopaminergic branches.[8] Figure 1 is a schematic representation of the simple cognitive deficit model of ADHD (column on the left) and a simplified account of the associated frontostriatal circuitry (column on the right). B, C, and S represent biology, cognition, and symptoms, respectively. The slashed C represents cognitive deficit; NE, norepinephrine; DA, dopamine; DLPFC, dorsolateral prefrontal cortex. ADHD results from impairments in the dopamine and norepinephrine systems, which play a crucial role in efficient communication in the brain's circuitry. The symptoms of ADHD are here considered to be caused by dysfunctions of neurocognitive control systems, more specifically, by deficits in inhibitory-based executive processes. While there is no consensus definition, "executive function" is usually taken to be a broad range of cognitive processes responsible for facilitating the pursuit of future goals and involved in the distribution of cognitive-energetic resources (i.e. effort) to activation

[6]"Pathogenetic models of ADHD have traditionally focused on molecules involved in neurotransmission and catecholamine synaptic dysfunction, including dopamine transporter DAT1 (SLC6A3), dopamine receptors DRD4, DRD5 and synaptosomal protein SNAP-25. More recently neural developmental genes including cadherin 13 (CDH13) and cGMP-dependent protein kinase I (PRKG1) have been associated with ADHD" (Cristino et al. 2014, p. 294; see also, Fowler et al. 2009; Sharp et al. 2009).

[7]Medication and behavioural interventions are difficult to isolate completely in order to compare their efficacy. In contexts in which pharmacological treatments are adopted as a consequence of the disorder being diagnosed (e.g. school and home), some form of behavioral intervention—more or less systematic—is usually implemented at the same time.

[8]See e.g. Barkeley (1997). All figures are from Sonuga-Barke (2005). Copyright © 2004 Society of Biological Psychiatry. Reprinted with permission of Elsevier.

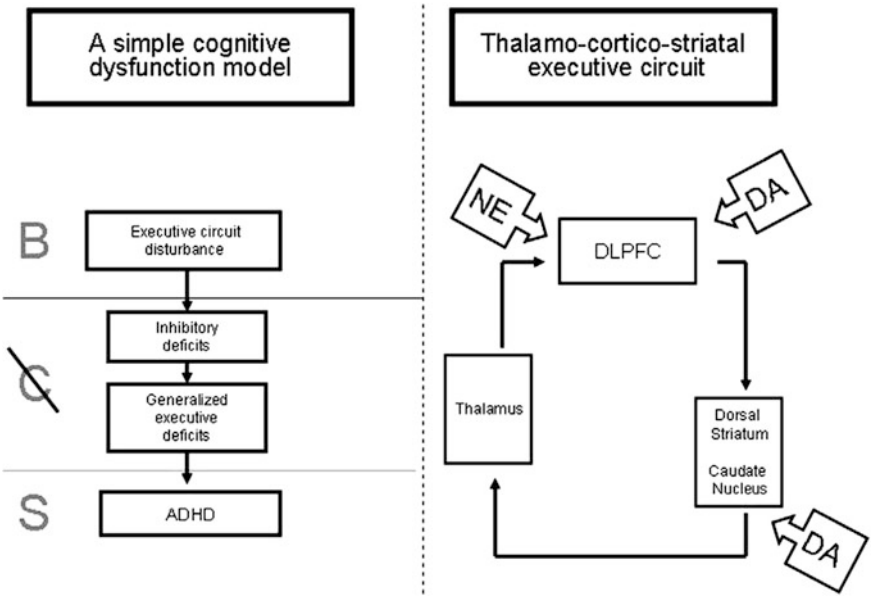


Fig. 1 Simple cognitive deficit model

and arousal systems to meet the changing demands of different situations.[9] At the neurobiologic level, "there is growing evidence that inhibitory control and other cognitive functions are underpinned by one of a family of […] basal ganglia-thalamocortical circuits […]. Data from structural and functional neuroimaging studies support the hypothesis that deficits in inhibitory-based executive functions in ADHD are associated with disturbances in this circuit […]. Dopamine […] is a key neuromodulator of this circuit" (Sonuga-Barke 2005, p. 1232).

According to a different simple causal model, the *motivational model*, ADHD results from impaired signaling of delayed rewards arising from disturbances in motivational processes, which involve frontoventral striatal reward circuits and mesolimbic branches terminating in the ventral striatum. Deficits are hence imputed to reward mechanisms, with ADHD being thought of as the outcome of "neurobiological impairment in the power and efficiency with which the contingency between present action and future reward is signaled. This leads to a reduction in the control exerted by future rewards on current behaviour" (Sonuga-Barke 2005,

[9]Examples of such processes include "planning and implementing strategies for performance, initiation and discontinuation of actions, inhibiting habitual or prepotent responses or task irrelevant information, performance monitoring, vigilant attention and set switching. Researchers have struggled to understand whether the broad range of 'executive' functions are supported by a single unitary process or a diverse array of cognitive processes" (Castellanos et al. 2006, p. 118). Current models are supported by neuroimaging and studies on focal lesions and tend to conceive executive function as a collection of higher-order cognitive control processes.

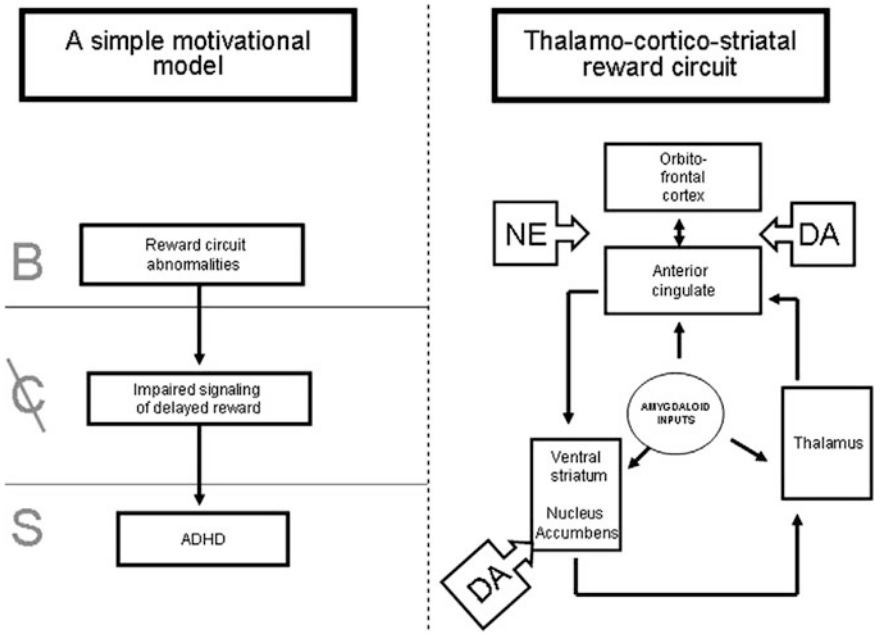


Fig. 2 Simple motivational model

p. 1233). This view of the disorder is supported by data on ADHD children's hypersensitivity to delay, difficulties in waiting for motivationally salient outcomes and in working effectively for prolonged periods of time. These are held to be related to alterations in another of the dopamine-modulated thalamocortical-basal ganglia circuits.[10] The circuit at stake here "links the ventral striatum (in particular, the nucleus accumbens) to frontal regions (especially the anterior cingulate and orbitofrontal cortex), connections that are reciprocated via the ventral pallidum and related structures through the thalamus. The amygdala also seems to be implicated in this system, possibly playing a role in defining the motivational significance of incentives" (Sonuga-Barke 2005, p. 1233). Figure 2 is a simple motivational model of attention-deficit/hyperactivity disorder (ADHD) (column on the left) and a simplified model of the frontostriatal circuitry involved (column on the right). Data suggest that the executive circuit considered in the first model presented above and the reward circuit considered here might *each* make *distinctive* contributions to the development of the disorder.

[10]"The fact that dopamine is a key neuro-modulator of both the executive and reward circuits therefore provides further support for the neurobiological plausibility for these cortico-basal ganglia models of AD/HD. At the same time, the fact that each circuit is influenced by inputs from different branches of the dopamine system confirms the differentiation of the pathways" (Sonuga-Barke 2003, p. 598).

These two models provide *simple* causal paradigms, in need of some implementation. Driven by the idea that theoretical models *combining* motivational and cognitive elements are needed, research is being carried out on the relations between these two models, and on the role of some *further* causal factors to be included. Different models have hence been advanced to shift from common simple deficits to *multiple neurodevelopmental pathway accounts*. The ultimate aim is to reach some *explanatory* account of the arrays of symptoms, by including, amongst others, the role of the social environment in shaping neurodevelopmental pathways to ADHD. The still elusive mechanisms connecting psychological activity and behavioral outcomes with low-level neurobiological processes are to be unraveled.

As a variation and extension of the motivational model, the *delay aversion* model has been put forward (Fig. 3), which stresses difficulties of ADHD-affected people in dealing with time-delayed rewards. Alterations in neurobiological circuits impair the signaling of delayed rewards, leading to impulsiveness; impulsiveness leads to failures to engage effectively with delay-rich environments; this failure to engage has the potential to elicit a negative punitive response from a parent or another significant adult (e.g. teacher), which, over time, leads to generalized delay aversion (column on the left). The failure to engage with delay-rich environments also constrains the experience of managing delay, and so reduces the opportunities to develop the organizational skills and strategies required to do this (column on the right). Delay aversion is expressed both as a compounding of existing impulsiveness and as a further elaboration of behavioral characteristics. Over time, various processes can reinforce a pattern of symptomatology and impairment can persist. For instance, negative parental responses elicited by the child's disorder, or uncertain and inconsistent environments in which future promised rewards or events were not delivered, place the child at risk of developing oppositionally. At the same time, the possibility that the child might *accommodate* the constraints

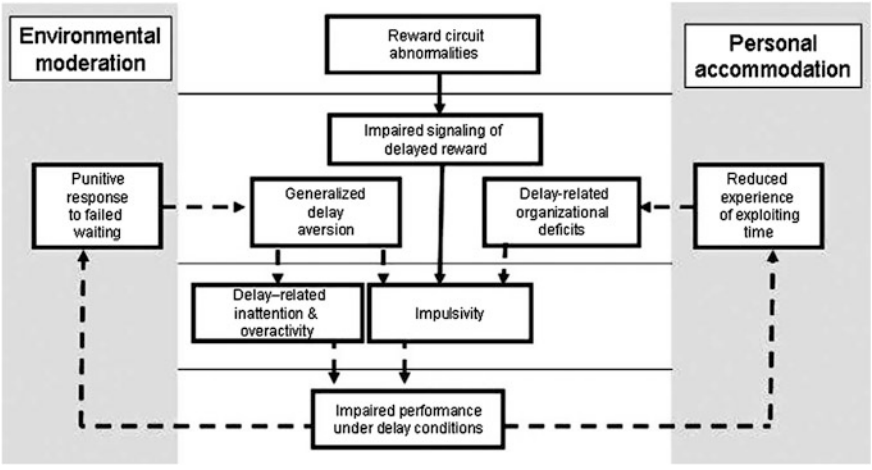


Fig. 3 Motivational developmental pathway model

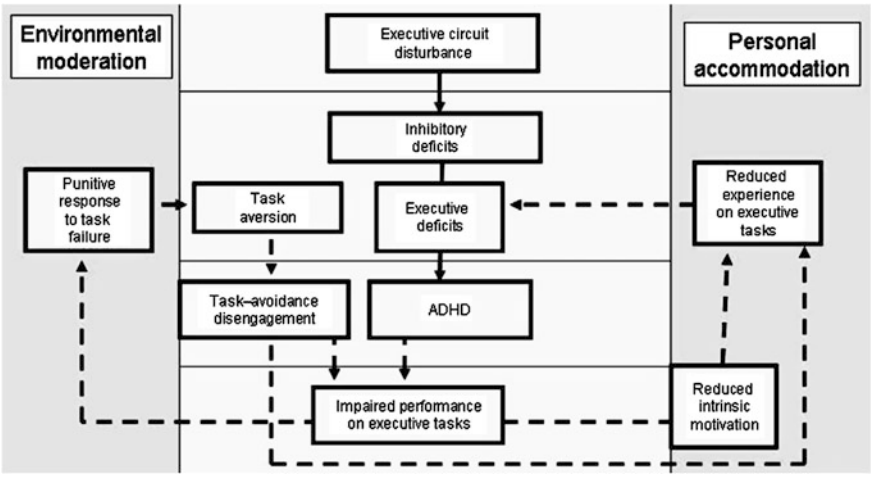


Fig. 4 Cognitive developmental pathway model

imposed by her underlying predisposition to impulsiveness and delay aversion should be considered. For instance, the hypothesis is being explored that ADHD children develop compensatory strategies to exploit limited processing time more effectively and to overcome deficits in working memory.

Similar considerations apply to a revision of the first model we considered, the cognitive deficit model. Figure 4 is a hypothetical cognitive developmental pathway model analogous to the motivational developmental pathway model. Negative/punitive responses might be elicited from significant adults (left column), potentially resulting in "executive-task aversion", which could in turn lead the child to avoid settings requiring executive effort and skills and early and premature disengagement. Reduced exposure to executive-type tasks might limit the opportunities to develop executive skills (right column). At the same time, failure on executive tasks might also reduce the extent to which tasks are intrinsically motivating. This will in turn reduce task effort and engagement in tasks, perpetuating the process.

The extended models are explicitly presented as hypothetical, and the extra elements included admittedly await further investigation. What is worth stressing here is that, with respect to the simple causal models, environmental and personal accommodation factors are explicitly added to the picture, with the aim of reaching a more accurate and articulated representation of the disorder. It is suggested that processes regulating the child's engagement with her environment and her developmentally significant experiences (e.g. educational agendas, cultural level to which one is exposed, a punitive parenting style, …) shape the course of her development.[11] Similarity should hold in developmental outcome within the same

[11]The most confirmed genes x environmental factors interactions are those between dopaminergic genes and maternal smoking, alcohol abuse during pregnancy and psycho-social adversity (as severe deprivation experience in early childhood).

community, with specific parenting styles and personal accommodation strategies being responsible for individual differences.

Much research is driven by the idea that "multiple pathway models may emerge as particularly powerful *explanatory tools* in this area" (Sonuga-Barke 2003, p. 597, italics added). Current investigations are directed, amongst others, towards a deeper understanding of how brain development in persons with ADHD differs from that in non-affected people of the same age, which roles emotions and motivation play in ADHD, which treatments tend to be more helpful and safer, and why and to what extent ADHD can overlap with other disorders. In the light of evidence that executive function deficits and delay aversion pathways are dissociated, but equally strongly associated with the combined ADHD symptomatology, some sort of integration of the two models is sought, by placing the pathways within a common neurobiological framework, and in the context of some understanding of the interplay between cortical and sub-cortical brain regions in the regulation of action, cognition and emotion/motivation. Evidence from, e.g., neuroanatomical, imaging, psychopharmacological, and clinical studies is collected to model cortical and subcortical interactions, in an attempt to elaborate multipathway, multilevel developmental accounts of the disorder incorporating different kinds of data and different neuropsychological effects.

Neuropathologic heterogeneity in ADHD might have major implications for the clinical management of the condition, and more articulated causal models may, in the long run, suggest ways to better tailor treatments. Pathways moderated by cultural or social factors are likely to be treated by psychosocial interventions, while pharmacology may focus on selective antagonists targeting specific dopamine circuits. "Overall, the field has witnessed notable progress as it converges on an understanding of ADHD in relation to disruption of a multicomponent self-regulatory system." (Nigg 2005, p. 1424). Research trends give some insight into the dynamics of modeling ADHD, through which accounts are progressively implemented and revised, building up different conceptions of the disorder and, hence, orienting different therapeutic practices.

3 Models of ADHD and Mechanistic Models: What Do They Explain, and How?

Which characteristics do these models exhibit, and which kinds of explanations do they provide? Some deep examination of this case study can stimulate reflections on behavioral and mechanical descriptions, and relations between them. To start with, *what* do we explain? One of the most interesting features with respect to accounts of such a relatively recently introduced pathology—whose definition, as we have seen, has varied—is the very *identification of the explanandum*, whose importance should not be underestimated and which clearly and significantly affects the elaboration of

any explanatory model. In the case of controversial disorders such as ADHD, we do not start from a single and shared description of a definitely isolated phenomenon. Craver states: "a mechanistic explanation must begin with an accurate and complete characterization of the phenomenon to be explained […]. To characterize the phenomenon correctly and completely is the first restrictive step in turning a model into an acceptable mechanistic explanation" (Craver 2006, pp. 368–369). In cases like the one at issue, the explanatory enterprise does *not* start from a single, accurate and complete description of the system under investigation. Rather, the identification of the disorder evolves along with the progressive identification of the relevant causal features and the elaboration of more and more comprehensive explanatory accounts. *Descriptions* of the disorder *and explanatory practice* can be thought of as *co-evolving*, with shifts on emphasis, taking predominantly to the foreground either cognitive deficit, motivational, environmental, or other aspects, or some combination of them.

Models are here aimed "to account for the *cardinal symptom domains* of impulsiveness, inattention, and hyperactivity; however a *number of other candidate-defining features clearly exist*" (Sonuga-Barke 2005, p. 1233, italics added), which can affect the explanatory account. The elaboration of the model hence starts from the *choice of a minimum set* of characterizing features that are taken to describe the disorder, and which will impinge on what will be identify as the explanans. While we struggle to unravel the causal mechanisms bringing about the disorder, characterizations of the disorders themselves are subject to revisions and social and historical factors, and change according to, e.g., ongoing research, discoveries and innovation. In dealing with psychiatric disorders, while we strive to identify some common network of causal mechanisms, isolating the explanandum system with precision is further complicated by a high rate of comorbidity. Specifically, ADHD can often co-occur with Tourette syndrome or rage attacks; Tourette syndrome, in turn, is often accompanied by depression and obsessive-compulsive disorder. The concurrence of different disorders makes it even more difficult to precisely draw the borders of the pathology to be modeled, disentangle the relevant variables, their mutual dependence, or dependence on some common cause, and detect their exact temporal sequence and interactions in time. The explanandum is some *provisional description* of the disorder, largely relying on what is taken to be the prevailing cluster of symptoms—which is, in turn, subject to change. Different kinds of symptoms have been associated with ADHD, have been attributed different diagnostic weight, and have all played a major role in the description and classification of the pathology. Not only can we agree that "as one incorporates more mechanistically relevant details into the model […], one correspondingly improves the quality of the explanandum" (Kaplan 2011, p. 347), but we can also observe that the re-description of the explanandum can result from an extension of the mechanistic account, and can in turn orientate the search for further mechanistically relevant details in some direction rather than others.

For the model to be explanatory, some mapping must occur between elements in the model and elements in the mechanism bringing about the explanandum.[12] The disease is modeled in the first place as a network of impaired normal mechanisms. To start with, deficits and malfunctions in specific steps of standard neurophysiological mechanisms are taken to explain the disorder which can in turn be deemed *sub-mechanisms* of ADHD.[13] The first two models considered above take different neurophysiological circuits to constitute the fundamental components of the disorder, and hence to provide the fundamental clues to understand it. However, further research suggests that the pathology can be adequately accounted for only as an articulated system including multiple, intertwined and heterogeneous levels. In the third and forth extended models, neurophysiological mechanisms are integrated with higher-level causal reinforcing factors, and deficits in the normal neurobiological functioning of the brain system are embedded in a wider characterization of the disorder, as *a broader pathological mechanism*. "Heuristically, the pathological mechanism and the causal lineage in which it becomes increasingly distant from normal behavior and increasingly global in its effects is the most relevant object of medical knowledge" (Nervi 2010, p. 219). A broadening of the causal picture can go hand in hand with a re-description of the disorder and the identification of what are taken as its distinctive features (e.g. the role of cognitive development). All these features obviously also affect clinical practice, insofar as modeling orientates therapeutic interventions at different—physiological and non-physiological—levels.

Instantiations of ADHD can vary significantly from one another, and what we model is something which will never occur as such: we are modeling the disease, while always encountering the diseased. What we are explaining is, in Murphy's terms, an "exemplar". On the one hand, mechanistic representations provide some manageable accounts of clinical conditions, overcoming the idiosyncrasies of the individual cases through the identification of some regular patterns expressed in the model. On the other hand, which model among a range of possible models will be referred to and the uses to which it will be put are dictated by some specific features of given instantiations of the disorder. When addressing ADHD, one of the difficulties is given by the need to account for the range of possible differences encountered in clinical practice,[14] due, for instance, to different person x environment interactions and different individual adaptation to developmental constraints. Some of the core elements the models presented include—those related to the

[12]See Kaplan and Craver (2011).

[13]On whether diseases are to be regarded as *pathological mechanisms*, conceived as separate and autonomous entities, or rather as *malfunctions* of physiological mechanisms, deemed as a *conceptual prior* over pathological mechanisms, see Nervi (2010), Moghaddam-Taaheri (2011).

[14]Let us recall that recovery rates can vary too. Even if it is commonly considered a childhood disorder, ADHD actually endures into adulthood in more than a half of the cases. The relevant interactions between environmental factors and neurodevelopmental components differ over time, according to age, and so do the symptoms. On ADHD in adulthood, see Karama and Evans (2013), Shaw et al. (2013).

involvement of dopaminergic circuits—are both crucial and quite unspecific, regarding the basic working of neurobiological underpinning mechanisms, holding for the general population of ADHD patients. Causal networks involved also include higher-level factors—such as parenting styles and social conditions—that, while affecting the whole the population in principle, might or might not make an *actual* difference in the single case, and can hence partly account for the degrees of variability, the possible arrays of outcome symptoms, and their different levels of severity. What is needed is some better understanding of the actions of the psychological and environmental factors on the neural underpinnings of the disorder, and some integration of subsystems—from changes in molecular, synaptic and cellular functions to sensory, cognitive and motor activities. This in turn requires integrating data from studies in a number of fields, like neuropsychology, developmental psychology, behavioral and cognitive neuroscience, and genetics.[15] ADHD cannot be modeled as a static medical condition, due to dysfunctions of a few isolated brain regions, but as a developmental trajectory involving different possible causes, mediators and outcomes.

So *how* are we explaining the disorder? Mechanistic modeling requires specification of the entities and activities operating in a system, and their *organization*: the behavior of the mechanism depends upon *what* the components *are and* on *how* the components and their activities are organized and interact with each other—bringing together, in our case, underlying neural and neurocognitive systems, neurodevelopmental processes, and environmental causal factors. While the behavioral description can rest content with the identification of the output behavior, the mechanical description requires some unraveling of the mechanism's working, and that is what the mechanist, explanatory model is called to specify. In discussing how *mechanical adequacy* must be assessed, Glennan (2005) believes the following questions should be addressed: Has the model identified *all* of the components in the mechanism? Have the components been *localized*? Does the model provide *quantitatively accurate* descriptions of the interactions and activities of each component? Does the model correctly represent the *spatial and temporal organization* of the mechanism? Models considered here are purely qualitative, and present different degrees of specification and graininess. To provide an adequate account of ADHD, the models start off at the neurobiological level, and then proceed "to build causal chains across intermediate cognitive or neuropsychologic and behavioral levels of analysis" (Sonuga-Barke 2005, p. 1232), moving up to environmental and social levels. Hence, a very fine-grained "zoomed-in" description of the neurophysiological mechanisms underpinning the disorder goes together with a "zoom out" at higher levels. Spatiotemporal scales are very different, and temporal features like the order and "rhythm" at which different level factors act are not specified. The inclusion of higher levels does not simply amount to situating the mechanistic system in its context, but to an attempt at *reshaping its very*

[15]See Halperin and Healey (2011), Coghill (2014) and Sonuga-Barke and Coghill (2014).

boundaries, which will not be clear-cut. The resulting mechanism is broader, and potentially open to further expansion.[16]

These aspects shed some light on the relations between behavioral and mechanical descriptions in the construction of an explanatory model of a disorder. Not only does the inclusion of, e.g., cultural, psychological and social factors provide a different explanatory framework, but also, and more in general, a different conception of the disorder, which cannot be easily and straightforwardly isolated. The extended models do not originate simply from some "looking around and up" after "looking down" (see Bechtel 2009),[17] to situate the mechanism by putting it into its proper context. They actually draw a different, more comprehensive system, re-define the boundaries of the mechanism, and thus draw a different disorder altogether. This has clear implications with respect to the localization of the system. Implicit in the classic disease model of mental disorders "is the assumption that mental disorders are discrete disease entities, […] which result from a dysfunction of neuropsychological/biologic mechanisms within the patient" (Sonuga-Barke 2005, p. 1231). Once mental disorders are taken as separate, discrete disease entities, "it is not surprising that much scientific psychopathology seems motivated by a quest to identify the *site* of the core dysfunctions that 'cause' the disorder" (Sonuga-Barke 2005, p. 1231, italics added). Broadening the range of potentially relevant causal factors like that in the extended models affects the very idea of *localization* of the target system: the system won't be isolated through some *spatial* localization, according to the "*site* of the core dysfunction", but rather in terms of the active role of relevant causal factors involved.[18]

The arrows in the figures represent the relations which are taken as explanatorily relevant. In general, etiopathogenetic causes are responsible for an initial increased vulnerability to mental disorders, and can include, e.g., genes, obstetric complications, urban birth and upbringing in extreme poverty, migrant status, chronic cannabis use, social isolation, and lack of support.[19] The activation of the pathophysiologial mechanisms—temporally more proximal—brings about the clinical manifestations and symptoms. The pathophysiological and/or etiopathogenetic factors figure in the explanation of the clinical features, and the clinical features provide a testing ground of the validity of the former.[20] Models at stake here provide *possible alternative explanations* of the disorder, *tentatively* accounting for the behavior of the whole

[16]On mechanistic modeling as an integrative and iterative process, see Boogerd et al. (2013).

[17]See also Bechtel (2010).

[18]McManus has convincingly argued that "too much emphasis has been given to the discreteness of parts, foreclosing the possibility that diffuse entities might be epistemically useful in the realm of mechanistic explanation" (McManus 2012, p. 532).

[19]Their common final effect grounding individual vulnerability to psychosis is supposed to consist in a sensitization of an individual's striatum, which is then expressed by modifications in dopamine release in the brain. Investigations are being carried on regarding both genetic and environmental etiologies of dopamine deficits and their effects on the dopamine system in early development.

[20]See Oulis (2013a, b).

system in terms of the working of its *pathophysiological mechanisms*, a complete account of the causal history of the mental disorder, including the etiopathogenetic factors, remaining an elusive goal. Our models support causal links by different kinds and amounts of evidence. While the causal relations assessed by relying on already accepted neurophysiological knowledge of standard functioning of the brain's processes are presented as—so-to-speak—"confirmed lines", reinforcing actions performed by higher-level causal components are admittedly presented as just hypothetical, in need of further specification. The working parts belong to different granularity levels, and relations between levels are to be clarified.

To count as genuinely explanatory, models do not necessarily need to include an awful lot of details. As recently stressed in some portions of the debate on abstraction and idealization in modeling,[21] in some cases less can be more, with the chosen level of abstraction and amount of details included to be evaluated always with respect to the context and goals of the investigation. "Building more of everything into the model does not automatically augment knowledge, and it can in fact obscure the situation by including details that are not relevant for producing the system property that is a particular project's focus" (O'Malley et al. 2014, p. 818). Adding more details does not per se yield a better explanation. On the one hand, the explanatory import is not to be measured on the basis of the number of details included, but, rather, on *which* details are included. On the other hand, some account of *organization* must count as a bar set on any *explanatory mechanistic* model *once the explanandum has been fixed*. While the context can dictate the level of fine-graininess and amount of details included from different levels described, some causal organization of the interacting multiple pathogenic factors must be specified for an understanding of how the constellation of typical symptoms are brought about, and in what respects ADHD differs from co-morbid as well as other psychiatric disorders. A thorough extended mechanistic explanation of the phenomenon should also include some clear indication of the intertwined operations of neurobiological alterations underpinning the relevant pathways and the psychological and environmental processes mediating and moderating them. The two extended models assess *that* environmental factors are causally relevant, but fail to explain *how* they act in contributing to the pathology functioning and reinforcing it.[22]

Organization "involves an internal division of causal labor whereby different components perform different causal roles […]. Given some effect or behavior, a system is organized with respect to that effect or behavior if (a) different components of the system make different contributions to the behavior, and (b) the components' differential contributions are *integrated*, exhibiting *specific interdependencies* (i.e. each component interacts in particular ways with a particular subset of other components)" (Levy and Bechtel 2013, pp. 243–244). The extended mechanistic explanatory models should hence clarify relations between what can be identified as different psychopathological sub-mechanisms within the same

[21]See Batterman (2009), Batterman and Rice (2014) and Rohwer and Rice (2013).

[22]Some open issues along these lines are stressed in Campbell et al. (2014).

disorder. Once environmental factors are brought into the picture as causally relevant for the neurodevelopmental disorder, if the model is meant to be *mechanistically explanatory*, interdependencies must be modeled as the key to the system's overall behavior. For instance, "how does a polymorphism in a (dopamine) risk-gene for ADHD translate into a neurobiological substrate and result in behaviors that warrant a diagnosis of ADHD in a developing child? (Durston and Konrad 2007, p. 374). *How* does the blocking of striatal dopamine transporters, achieved through pharmacological treatment, alleviate behavioral symptoms? *How* does an increase in striatal activation improve motor function? Integrated knowledge of levels is likely to make us gain in control over the disorder,[23] which is definitely one of the major aims of neuropsychiatry, and to shed some light on other neurodevelopmental disorders as well.

Returning to Glennan's requirements, what sense does it make to talk of the adequacy of an explanatory model in this kind of context? The adequacy of the model has to do with both the identification of genuinely causal relations between variables and their organization, with their mutual constraints, and with what the explanatory model is going to be *used for*. Different, not incompatible, explanations, with different levels of graininess can be provided for one and the same phenomenon. The issue here is not whether, for instance, impulsivity or aggressiveness involve the brain, but whether a neural or neuropsychologic or social level of analyses "is *the most useful level* for understanding why this disorder develops" (Nigg et al. 2005, p. 1224, italics added). The field in which the investigation is pursued, its methods and purpose shape the kind of questions raised, the methodological and conceptual tools employed to answer them, and the sort of answers accepted. Neurobiologic and biochemical activity may not be the most adequate level to focus on if—for instance—the purpose is providing an adequate explanatory model to support some behavioral interventions on ADHD in a school context. Vice versa, the impact of parenting styles on neurodevelopmental features might prove less interesting in the search for increasingly effective pharmacologic medication on behalf of some pharmaceutical company. If it can be agreed that "the way in which mechanisms are investigated shapes the kinds of explanations in which those mechanisms figure" (Andersen 2014, p. 276), it equally holds that the use to which explanations are to be put and the contexts in which they are elaborated shape the search for the mechanisms they are supposed to exhibit, and how mechanistic systems are isolated and described.

Concluding, in what respects do the models considered provide explanatory accounts of the neuropsychiatric disorder at stake in mechanistic terms, and to what extent do they fulfill neo-mechanist desiderata? If we reflect on the dynamics of modeling *in scientific practice* in contexts like the one at issue, features like an accurate and complete description of the explanandum, and the specification of interactions and organization seem to be seen at most as a regulatory ideal.

[23]On the relation between mechanistic knowledge and control, see Craver and Darden (2013, Chap. 11).

The mechanistic models described here deal with a disorder lying at the crossroad of systems which are known to various extents. Resulting models look as a matter of fact like some *combination* of mechanism sketches and mechanism schemata, or like some set of how-possibly, how-plausibly and how-actually models of sub-systems, to be integrated through the merging of knowledge from different fields of enquiry, at different levels of graininess. At the time that a given model is proposed as a representation of the working of the disorder, we might not know exactly to what extent parts of our model are approximations and distortions, and which parts represent more or less accurately the actual features of the disorder—whose description, on top of that, is in turn subject to changes and revisions. The plausibility of the model, the possibility or need for revisions must be evaluated in the long run. The mechanistic models considered allow us to answer some range of why questions on the disorder, are admittedly incomplete and partly hypothetical and constructed piecewise, with poor cues on modes of interaction between causal factors.

Any progress in mechanistic understanding of some level further constrains the space of possible mechanisms underpinning the disorder. In the process, one obtains a more accurate identification and description of the mental dysfunction which, in turn, enables a more accurate identification of what produces it, with the relational properties of models and target systems being produced simultaneously.[24] The description of the underlying mechanism can reshape the definition of the disease, which can vary and, in turn, orientate further search for explanatory accounts. Disciplinary contextualization, and unraveling of underpinning assumptions depending on disciplinary standpoints, might help shed light on relationships between different approaches and tentative models. Interdependence must be brought into light between the different partial explanatory accounts of what can be drawn as sub-systems if some grasping of the causal complexity exhibited by mutually reinforcing networks of causal mechanisms (see Kendler et al. 2011) is to be reached. The target of a coherent, comprehensive and integrated model of ADHD might be quite far off, neuropsychiatry as a field being, with respect to mechanistic stances, in many senses still "at the 'how possibly' stage" (Kendler 2008a, b, p. 899).

References

Andersen, H.: A field guide to mechanisms: Part I. Philos. Compass **9**(4), 274–283 (2014)
Barkeley, R.A.: ADHD and the Nature of Self-Control. Guilford Press, New York (1997)
Batterman, R.: Idealization and modeling. Synthese **169**, 427–446 (2009)
Batterman, R., Rice, C.: Minimal model explanations. Philos. Sci. **81**, 349–376 (2014)

[24]On the co-construction of models and their target systems, see Knuuttila and Boon (2011), Green (2013).

Bechtel, W.: Looking down, around, and up: mechanistic explanation in psychology. Philos. Psychol. **22**, 543–564 (2009)
Bechtel, W.: The downs and ups of mechanistic research: circadian rhythm as an exemplar. Erkenntnis **73**, 313–328 (2010)
Boogerd, F.C., Bruggeman, F.J., Richardson, R.C.: Mechanistic explanations and models in molecular systems biology. Found. Sci. **18**, 725–744 (2013)
Campaner, R.: Explanatory pluralism in psychiatry: what are we pluralists about, and why? In: Galavotti, M.C., et al. (eds.) New Directions in the Philosophy of Science, pp. 87–103. Springer, Dordrecht (2014)
Campaner, R.: The interventionist theory and mental disorders. In: Gonzalez, W. (ed.) Causal Explanation and Philosophy of Psychology: New Reflections on James Woodward's Contribution. Springer, Dordrecht (forthcoming)
Campbell, S.B., Halperin, J.M., Sonuga-Barke, E.J.: A developmental perspective on attention-deficit/hyperactivity disorder. In: Lewis, M., Rudolph, K.D. (eds.) Handbook of Developmental Psychopathology, pp. 427–448. Springer, New York (2014)
Castellanos, F.X., et al.: Characterizing cognition in ADHD: beyond executive dysfunction. TRENDS Cognit. Sci. **10**, 117–123 (2006)
Coghill, D.: Editorial: acknowledging complexity and heterogeneity in causality. Implications of recent insights into neuropsychology of childhood disorders for clinical practices. J. Child Psychol. Psychiatry **55**, 737–740 (2014)
Craver, C.: When mechanistic models explain. Synthese **153**, 355–376 (2006)
Craver, C., Darden, L.: In Search of Mechanisms. The University of Chicago Press, Chicago (2013)
Cristino, A.S., et al.: Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. Mol. Psychiatry **19**, 294–301 (2014)
Durston, D., Konrad, K.: Inegrating genetic, psychopharmacological and neuroimaging studies: a converging methods approach to understanding the neurobioogy of ADHD. Dev. Rev. **27**, 374–395 (2007)
Fowler, T., et al.: Psychopathy trait scores in adolescents with childhood ADHD: the contribution of genotypes affecting MAOA, 5HTT and COMT activity. Psychiatr. Genet. **19**(6), 312–319 (2009)
Glennan, S.: Modeling Mechanisms. Stud. Hist. Philos. Biol. Biomed. Sci. **36**, 443–464 (2005)
Green, S.: When one model is not enough: combining epistemic tools in systems biology. Stud. Hist. Philos. Biol. Biomed. Sci. **44**, 170–180 (2013)
Halperin, J., Healey, D.: The influence of environmental enrichment, cognitive enhancement, and physical exercise on brain development: can we alter the developmental trajectory of ADHD? Neurosci. Behav. Rev. **35**, 621–634 (2011)
Kaplan, D.: Explanation and description in computational neuroscience. Synthese **183**, 339–373 (2011)
Kaplan, D., Craver, C.: The explanatory force of dynamical and mathematical models in neuroscience: a mechanistic perspective. Philos. Sci. **78**, 601–627 (2011)
Karama, S., Evans, A.: Neural correlates of ADHD in adulthood. Biol. Psychiatry **74**, 558–559 (2013)
Kendler, K.: Toward a philosophical structure for psychiatry. Am. J. Psychiatry **162**, 433–440 (2005)
Kendler, K.: Explanatory models for psychiatric illness. Am. J. Psychiatry **165**, 695–702 (2008a)
Kendler, K.: Review of Carl Craver's "Explaining the Brain". Psychol. Med. **38**, 899–901 (2008b)
Kendler, K.: Levels of explanation in psychiatric and substance use disorders: implications for the development of an etiologically based nosology. Molecul. Psychiatry **17**, 1–18 (2012)
Kendler, K., Zachar, P., Craver, C.: What kinds of things are psychiatric disorders? Psychol. Med. **41**, 1143–1150 (2011)
Knuuttila, T., Boon, M.: How do models give us knowledge? The case of Carnot's ideal heat engine. Eur. J. Philos. Sci. **1**, 309–334 (2011)

Levy, A., Bechtel, W.: Abstraction and the organization of mechanisms. Philos. Sci. **80**, 241–261 (2013)
Machamer, P., Darden, L., Craver, C.F.: Thinking about mechanisms. Philos. Sci. **67**, 1–25 (2000)
McManus, F.: Development and mechanistic explanation. Stud. Hist. Philos. Biol. Biomed. Sci. **43**, 532–541 (2012)
Moghaddam-Taaheri, S.: Understanding pathology in the context of physiological mechanisms: the practicality of a broken-normal view. Biol. Philos. **26**, 603–611 (2011)
Murphy, D.: Explanation in psychiatry. Philos. Compass **5**(7), 602–610 (2010)
Murphy, D.: Conceptual foundations of biological psychiatry. In: Gifford, F. (ed.) Philosophy of Medicine, pp. 425–451. Elsevier, Amsterdam (2011)
Nervi, M.: Mechanisms, malfunctions and explanation in medicine. Biol. Philos. **25**, 215–228 (2010)
Nigg, J.T.: Neuropsychologic theory and findings in attention-deficit/hyperactivity disorder: the state of the field and salient challenges for the coming decade. Biol. Psychiatry **57**, 1424–1435 (2005)
Nigg, J.T., Willcutt, E.G., Doyle, A.E., Sonuka-Barke, J.S.: Causal heterogeneity in attention-deficit/hyperactivity disorder: do we need neuropsychologically impaired subtypes? Biol. Psychiatry **57**, 1224–1230 (2005)
O'Malley, M., et al.: Multilevel research strategies and biological systems. Philos. Sci. **81**, 811–828 (2014)
Oulis, P.: Toward a unified methodological framework for the science and practice of integrative psychiatry. Philos. Psychiatry Psychol. **20**, 113–126 (2013a)
Oulis, P.: Explanatory coherence, partial truth and diagnostic validity in psychiatry. In Karakostas, V., Dieks, D. (eds.) EPSA11 Perspectives and Foundational Problems in Philosophy of Science, pp. 429–440. Springer, Dordrecht (2013b)
Rohwer, Y., Rice, C.: Hypothetical pattern idealization and explanatory models. Philos. Sci. **80**, 334–355 (2013)
Salmon, W.: Scientific Explanation and the Causal Structure of the World. Princeton University Press, Princeton (1984)
Salmon, W.: Causality and Explanation. Oxford University Press, New York (1998)
Schaffner, K.: Reduction and reductionism in psychiatry. In: Fulford, K.W.M., et al. (eds.) The Oxford Handbook of Philosophy and Psychiatry, pp. 1003–1022. Oxford University Press, Oxford (2013)
Sharp, S.I., McQuillin, A., Gurling, H.: Genetics of attention-deficit hyperactivity disorder (ADHD). Neuropharmacology **57**, 590–600 (2009)
Shaw, P., et al.: Trajectories of cerebral cortical developmental in childhood and adolescence and adult ADHD disorder. Biol. Psychiatry **74**, 599–606 (2013)
Sirgiovanni, E.: The mechanistic approach to psychiatric classification. Dialogues Philos. Mental Neuro Sci. **2**, 45–49 (2009)
Sonuga-Barke, E.J.: The dual pathways model of AD/HD: an elaboration of neuro-developmental characteristics. Neurosci. Behav. Rev. **27**, 593–604 (2003)
Sonuga-Barke, E.J.: Causal models of attention-deficit/hyperactivity disorder: from common simple deficits to multiple developmental pathways. Biol. Psychiatry **57**, 1231–1238 (2005)
Sonuga-Barke, E.J., Coghill, D.: Introduction: the foundations of next generation attention-deficit hyperactivity disorder neuropsychology: building on progress during the last 30 years. J. Child Psychol. Psychiatry **55**, 1–5 (2014)

Chaos and Stochastic Models in Physics: Ontic and Epistemic Aspects

Sergio Caprara and Angelo Vulpiani

Abstract There is a persistent confusion about determinism and predictability. In spite of the opinions of some eminent philosophers (e.g., Popper), it is possible to understand that the two concepts are completely unrelated. In few words we can say that determinism is ontic and has to do with how Nature behaves, while predictability is epistemic and is related to what the human beings are able to compute. An analysis of the Lyapunov exponents and the Kolmogorov-Sinai entropy shows how deterministic chaos, although with an epistemic character, is non subjective at all. This should clarify the role and content of stochastic models in the description of the physical world.

1 Introduction

In the last decades scientists and philosophers showed an intense interest for chaos, chance and predictability. Some aspects of such topics are rather subtle, and in the literature is not unusual to find wrong statements. In particular it is important to avoid confusion on the fact that to be deterministic (or stochastic) is an ontic property of a system, i.e. related to its own nature independently of our knowledge; while predictability, and somehow chaos, have an epistemic character, i.e. depend on our knowledge. We will see how the introduction of a probabilistic approach in deterministic chaotic systems, although with an epistemic character, is not subjective.

Often in the past, the central goal of science has been though to be "prediction and control", we can mention von Neumann's belief that powerful computers and a clever use of numerical analysis would eventually lead to accurate forecasts, and even to the control, of weather and climate:

S. Caprara (✉) · A. Vulpiani
Dipartimento di Fisica, Università Sapienza, Roma, Italy
e-mail: sergio.caprara@roma1.infn.it

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_8

The computer will enable us to divide the atmosphere at any moment into stable regions and unstable regions. Stable regions we can predict. Unstable regions we can control.[1]

The great scientist von Neumann was wrong, but he did not know the phenomenon of deterministic chaos.

About half a century ago, thanks to the contribution of M. Hénon, E. Lorenz (see e.g., Lorenz 1963) and B.V. Chirikov (to cite just some of the most eminent scientists in the field), deterministic chaos was (re)discovered. Such an event sure was scientifically important, e.g., as it clarifies topics like the different possible origins of the statistical laws and the intrinsic practical limits of the predictions. On the other hand, one has to admit that the term "deterministic chaos" can be seen as an oxymoron and induced the persistence of a certain confusion about concepts as determinism, predictability and stochastic laws. Our aim is to try to put some order into this matter, discussing some aspects of deterministic chaos which, in our opinion, are often misunderstood, leading to scientifically, as well as philosophically, questionable and confused claims.

In spite of the fact that it is quite evident that Maxwell, Duhem, and Poincaré (see e.g., Poincaré 1892) understood in a clear way the distinction between determinism and chaos, in the recent literature one can find a large spectrum of wrong statements on the conceptual impact of deterministic chaos, see Campbell and Garnett (1882). For instance, Prigogine and Stengers (1994) claim that *the notion of chaos leads us to rethink the notion of "law of nature"*. In a book on statistical physics (Vauclair 1993), one can read that as consequence of chaos *the deterministic approach fails*. Sir James Lighthill (1986) in a lecture to the Royal Society on the 300th anniversary of Newton's Principia shows how to confuse determinism and prediction: *We are all deeply conscious today that the enthusiasm of our forebears for the marvelous achievements of Newtonian mechanics led them to make generalization in this area of predictability, which indeed we may generally have tended to believe before 1960, but which we now recognize were false. We collectively wish to apologize for having misled the generally educated public by spreading ideas about the determinism of systems satisfying Newton's laws of motion, that after 1960 were to be proved incorrect*.

Chaos presents both ontic and epistemic aspects[2] which may generate confusion about the real conceptual relevance of chaos. We shall see that chaos allows us to unambiguously introduce probabilistic concepts in a deterministic world. Such a possibility is not merely the consequence of our limited knowledge of the state of the system of interest. Indeed, in order to account for this limited knowledge, one usually relies on a coarse-grained description, which requires a probabilistic approach. We will see that many important features of the dynamics do not depend on the scale ϵ of the graining, if it is fine enough. At the same time, many results for

[1]Cited by Dyson (2009).

[2]We shall see how determinism refers to ontic descriptions, while predictability (and, in some sense, chaos) has an epistemic nature.

the $\epsilon \to 0$ limit do not apply to the cases with $\epsilon = 0$. Therefore, the probabilistic description of chaotic systems reveals one more instance of singular limits.

2 About Determinism

The word determinism has often been used in fields other than physics, such as psychology and sociology, causing some bewilderment. There have been some misunderstandings about the meaning of determinism, and because, at times, determinism has been improperly associated with reductionism, mechanicism and predictability (Chibbaro et al. 2014), it seems to us that a brief review of the notion of determinism is not useless.

For example, unlike the majority of modern physicists and mathematicians, by deterministic system Popper (1992) means a system governed by a deterministic evolution law, whose evolution can be in principle predicted with arbitrary accuracy:

Scientific determinism is the doctrine that the state of any closed physical system at any future instant can be predicted.

In other words, Popper confuses determinism and prediction.

On the contrary, Russell gives the following definition, which is in agreement with the present mathematical terminology:

A system is said to be "deterministic" when, given certain data $e_1, e_2, \ldots, e_n$ *at times* $t_1, t_2, \ldots, t_n$*, respectively, concerning this system, if* E_t *is the state of the system at any (later) time t, there is a functional relation of the form*

$$E_t = f(e_1, t_1, e_2, t_2, \ldots, e_n, t_n).$$

In the definition of Russell practical prediction is not mentioned.

The confusion about determinism and predictability is not isolated, see, e.g., Stone (1989) and Boyd (1972) who examine in great detail arguments about the widespread opinion that *human behavior is not deterministic because it is not predictable*.

Determinism amounts to the metaphysical doctrine that same events always follow from same antecedents. But, as Maxwell had already pointed out in 1873, it is impossible to confirm this fact, because nobody has ever experienced the same situation twice:

It is a metaphysical doctrine that from the same antecedents follow the same consequences. No one can gainsay this. But it is not of much use in a world like this, in which the same antecedents never again concur, and nothing ever happens twice … The physical axiom which has a somewhat similar aspect is "that from like antecedents follow like consequences". But here we have passed … from absolute accuracy to a more or less rough approximation.

In these few lines, Maxwell touches on issues which will be later investigated, and anticipates their solution. The issues are:

1. the impossibility of proving (or refuting) the deterministic character of the laws of Nature;
2. the practical impossibility of making long-term predictions for a class of phenomena, referred to here as chaotic, despite their deterministic nature.

After the development of quantum mechanics, many think that discussing the deterministic nature of the laws of physics is too academic an exercise to deserve serious consideration. For instance, in a speech motivated by the heated controversy on chaos and determinism between philosophers and scientists, (van Kampen 1991) bluntly said that the problem does not exist, as it is possible to show that:

the ontological determinism à la Laplace can neither be proved nor disproved on the basis of observations.[3]

It is not difficult to realize that determinism and predictability constitute two quite distinct issues, and the former does not imply the latter. Roughly speaking, determinism can be traced back to a vision of the nature of causality and can be cast in mathematical terms, by saying that the laws of nature are expressed by ordinary (or partial) differential equations. However, as noted by Maxwell, the objectively ontological determinism of the laws of nature cannot be proven; but one might find it convenient to use deterministic descriptions. Moreover, even at a macroscopic level, many phenomena are chaotic and, in some sense, appear to be "random". On the other hand, the microscopic phenomena described by quantum mechanics, fall directly within a probabilistic framework. When referring to observable properties, they appear ontologically and epistemologically non-deterministic.

3 Two Explicit Examples

In order to clarify the concepts of determinism, predictability and chaos let us discuss two deterministic systems whose behaviors are rather different. They do not have particular own relevance, their choice is motivated just for pedagogical reasons:

Example A The pendulum (of length L):

$$\frac{d^2\theta}{dt^2} = -\frac{g}{L}\sin\theta. \tag{1}$$

According to well known mathematical theorems on differential equations the following results hold:

[3]In brief, van Kampens argument is the following. Suppose the existence of a world A which is not deterministic and consider a second world B obtained from the first using the following deterministic rule: every event in B is the copy of an event occurred one million years earlier in A. Therefore, all the observers in B and their prototypes live the same experiences despite the different natures of the two worlds.

(a) the initial condition $(\theta(0), d\theta(0)/dt)$ determines in a unique way the state of the system $(\theta(t), d\theta(t)/dt)$ at any time t, in other words the system is deterministic;
(b) the motion is periodic, i.e., there exists a time T (depending on the initial conditions) such that

$$\left(\theta(t+T), \frac{d\theta(t+T)}{dt}\right) = \left(\theta(t), \frac{d\theta(t)}{dt}\right);$$

(c) the time evolution can be expressed via a function $F(t, \theta(0), d\theta(0)/dt)$:

$$\theta(t) = F\left(t, \theta(0), \frac{d\theta(0)}{dt}\right).$$

The function F can be explicitly written only if $\theta(0)$ and $d\theta(0)/dt$ are small (and, in such a case, $T = 2\pi\sqrt{L/g}$ is a constant, independent of the initial conditions); however, in the generic case, F can be easily determined with the desired precision.

Example B Bernoulli's shift:

$$x_{t+1} = 2x_t \ mod\, 1. \tag{2}$$

Where the operation *mod* 1 corresponds to taking the fractional part of a number, e.g., $1.473 \ mod\, 1 = 0.473$. It is easy to understand that the above system is deterministic: x_0 determines x_1, which determines x_2 and so on. Let us show that the above system is chaotic: a small error in the initial conditions doubles at every step. Suppose that x_0 is a real number in the interval $[0, 1]$, it can be expressed by an infinite sequence of 0 and 1:

$$x_0 = \frac{a_1}{2} + \frac{a_2}{4} + \ldots + \frac{a_n}{2^n} + \ldots,$$

where every a_n takes either the value 0 or the value 1. The above binary notation allows us to determine the time evolution by means of a very simple rule: at every step, one has just move the "binary point" of the binary expansion of x_0 by one position to the right and eliminate the integer part. For example, take

$$x_0 = 0.11010000101110101010101100\ldots$$

$$x_1 = 0.10100001011101010101011100\ldots$$

$$x_2 = 0.01000010111010101010110 0\ldots$$

$$x_3 = 0.10000101110101010101100\ldots$$

and so on. In terms of the sequence $\{a_1, a_2, \ldots\}$, it becomes quite clear how crucially the temporal evolution depends on the initial condition. Let us consider two initial conditions $x_0^{(1)}$ and $x_0^{(2)}$ such that $|x_0^{(1)} - x_0^{(2)}| < 2^{-M}$ for some arbitrary (large) integer number M, this means that $x_0^{(1)}$ and $x_0^{(2)}$ have the first M binary digits identical, and they may differ only afterwards. The above discussion shows that the distance between the points increases rapidly: for $t < M$ one has an exponential growth of the distance between the two trajectories

$$|x_t^{(1)} - x_t^{(2)}| \sim |x_0^{(1)} - x_0^{(2)}| 2^t .$$

As soon as $t > M$, one can only conclude that $|x_t^{(1)} - x_t^{(2)}| < 1$. Our system is chaotic: even an arbitrarily small error in the initial conditions eventually dominates the dynamics of the system, making long-term prediction impossible.

From the above discussion we saw how in deterministic systems one can have the following possible cases (in decreasing order of predictability):

(i) Explicit possibility to determine the future (pendulum in the limit of small oscillations);
(ii) Good control of the prediction, without an explicit solution (pendulum with large oscillations);
(iii) Chaos and practical impossibility of predictability (Bernoulli's shift).

3.1 About the Ontic/Epistemic Character of Chaos

One should also beware of the possible confusion between ontic and epistemic descriptions, when studying the topic of chaos. Determinism simply means that: given the same initial state $\mathrm{X}(0)$, one always finds the same evolved state $\mathrm{X}(t)$, at any later time $t > 0$. Therefore, determinism refers exclusively to ontic descriptions, and it does not deal with prediction. This has been clearly stressed by Atmanspacher (2002), in a paper by the rather eloquent title *Determinism is ontic, determinability is epistemic*. This distinction between ontic and epistemic descriptions was obvious to Maxwell; after having noted the metaphysical nature of the problem of determinism in physics, he stated that:

There are certain classes of phenomena ... in which a small error in the data only introduces a small error in the result ... There are other classes of phenomena which are more complicated, and in which cases of instability may occur.

Also for Poincaré the distinction between determinism and prediction was rather clear, on the contrary, Popper (1992) confused determinism and prediction.

4 Chaos and Asymptotics

Here, we briefly recall the essential properties of a deterministic chaotic system:

(i) The evolution is given by a deterministic rule, for example, by a set of differential equations;

(ii) Solutions sensitively depend on the initial conditions: i.e., two initially almost identical states $\mathrm{X}(0)$ and $\mathrm{X}'(0)$, with a very small initial displacement $|\mathrm{X}'(0) - \mathrm{X}(0)| = \delta_0$, become separated at an exponential rate:

$$|\mathrm{X}'(t) - \mathrm{X}(t)| = \delta_t \sim \delta_0 \, e^{\lambda t}, \tag{3}$$

where λ is positive and is called the Lyapunov exponent, for Bernoulli's shift $\lambda = \ln 2$;

(iii) The evolution of the state $\mathrm{X}(t)$ is not periodic and appears quite irregular, similar in many respects to that of random systems.

The sensitive dependence on the initial condition drastically limits the possibility of making predictions: if the initial state is known with a certain uncertainty δ_0, the evolution of the system can be accurately predicted with precision Δ only up to a time that depends on the Lyapunov exponent. This quantity is inherent in the system and does not depend on our ability to determine the initial state; hence, recalling Eq. (3), the time within which the error on the prediction does not exceed the desired tolerance is:

$$T_p \sim \frac{1}{\lambda} \ln \frac{\Delta}{\delta_0}. \tag{4}$$

The sensitivity to initial conditions introduces an error in predictions which grows exponentially in time. As the Lyapunov exponent λ is an intrinsic characteristic of the system, predictions remain meaningful only within a time given by Eq. (4); therefore, it is well evident that a deterministic nature does not imply the possibility of an arbitrarily accurate prediction.

Let us note that, since X is in a bounded domain, some accuracy is needed in the definition of the Lyapunov exponent: before one has to take the limit $\delta_0 \to 0$ and then $t \to \infty$:

$$\lambda = \lim_{t \to \infty} \lim_{\delta_0 \to 0} \frac{1}{t} \ln \left(\frac{\delta_t}{\delta_0} \right).$$

Another important characterisation of the dynamics is given by the Kolmogorov-Sinai entropy, h_{KS}, defined as follows. Just for the sake of simplicity we consider a system with discrete time: let $\mathcal{A} = \{A_1, \ldots, A_N\}$ be a finite partition

of the phase space (the space of configurations of a given system under study), made up of the N disjoint sets A_i, and consider the sequence of points

$$\{\mathrm{x}_1, \ldots, \mathrm{x}_n, \ldots\},$$

which constitutes the trajectory with initial condition x_0. This trajectory can be associated with the symbol sequence

$$\{i_0, i_1, \ldots, i_n, \ldots\}, \tag{5}$$

where $i_k = j$ if $\mathrm{x}_k \in A_j$.

Once a partition $\mathcal{A}$ has been introduced, the coarse-grained properties of chaotic trajectories can be therefore studied through the discrete time sequence (5). Let $C_m = (i_1, i_2, \ldots i_m)$ be a "word" (a string of symbols) of length m and probability $P(C_m)$. The quantity

$$H_m = \sum_{C_m} P(C_m) \ln P(C_m) \tag{6}$$

is called the block entropy of the m-sequences.[4] In the limit of infinitely long sequences, the asymptotic entropy increment

$$h_S(\mathcal{A}) = \lim_{m \to \infty} (H_{m+1} - H_m)$$

is called the Shannon entropy, and in general depends on the partition $\mathcal{A}$. Taking the largest value over all possible partitions we obtain the so-called Kolmogorov-Sinai entropy:

$$h_{KS} = \sup_{\mathcal{A}} h_S(\mathcal{A}).$$

A more intuitive definition of h_{KS} starts from the partition $\mathcal{A}_\epsilon$ made of a grid of hypercubes with sides of length ϵ, and takes the following limit:

$$h_{KS} = \lim_{\epsilon \to 0} h(\epsilon),$$

where $h(\epsilon) = h_S(\mathcal{A}_\epsilon)$.

Naively, one might consider chaos in deterministic systems to be illusory, just a consequence of our observational limitations. Apparently, such a conclusion is confirmed by the fact that important measures of the dynamical complexity, such as the Lyapunov exponent λ and the Kolmogorov-Sinai entropy h_{KS}, are defined via finite, albeit arbitrarily high, resolutions. For instance, in the computation of λ one

[4] Shannon (1948) showed that, once the probabilities $P(C_m)$ are known, the entropy (6) is the unique quantity which measures, under natural conditions, the surprise or information carried by $\{C_m\}$.

considers two trajectories, which are initially very close $|\mathrm{X}(0) - \mathrm{X}'(0)| = \delta_0$ and diverge in time from each other. Similarly, h_{KS} is computed introducing a partition of the phase space, whose elementary cells have a finite size ϵ. However, in the small-ϵ limit, $h(\epsilon)$ asymptotically tends to a value (h_{KS}) that no longer depends on ϵ, as happens to λ in the small-δ_0 limit. Therefore, λ and h_{KS} can be considered intrinsic properties of the dynamics themselves: they do not depend on our observational ability, provided it is finite, i.e., provided ϵ and δ_0 do not vanish. According to Primas (2002), measures of stability, such as the Lyapunov exponent, concern ontic descriptions, whereas measures of information content or information loss, such as the Kolmogorov-Sinai entropy, relate to epistemic descriptions. We agree as far as stability is concerned. Regarding the epistemic character of h_{KS}, we observe that the Shannon entropy of a sequence of data, as well as the Kolmogorov-Sinai entropy, enjoy an epistemic status from a certain viewpoint, but not from another. The epistemic status arises from the fact that information theory deals with transmission and reception of data, which is necessarily finite. On the other hand, h_{KS} is definitely an objective quantity, which does not depend on our observational limitations, as demonstrated by the fact that it can be expressed in terms of Lyapunov exponents (Cencini et al. 2009). We note that the ϵ-entropy $h(\epsilon)$ can be introduced even for stochastic processes, therefore it is a concept which links deterministic and stochastic descriptions.

5 Chaos and Probability

After the (re)discovery of chaos in deterministic systems, owing to the presence of irregular and unpredictable behaviours, it is quite natural to adopt a probabilistic approach even in the deterministic realm. Let us assume that we known the probability density of configurations in phase space at the initial time $\rho(\mathrm{x}, 0)$, it is possible to write down its time evolution law:

$$\rho(\mathrm{x}, 0) \rightarrow \rho(\mathrm{x}, t)\,.$$

Under certain conditions (mixing[5]) one has that al large time the probability density approaches a function which does not depend on $\rho(\mathrm{x}, 0)$:

$$\lim_{t\to\infty} \rho(\mathrm{x}, t) = \rho^{inv}(\mathrm{x})\,, \tag{7}$$

[5]The precise definition of mixing in dynamical systems requires several specifications and technicalities. To have an idea, imagine to put flour and sugar, in a given proportion (say 40 and 60 %, respectively) and initially separated, in a jar with a lid. After shaking the jar for a sufficiently long time, we expect the two components to be *mixed*, i.e., the probability to find flour or sugar in every part of the jar matches the initial proportion of the two components: a teaspoonful of the mixture taken at random will contain 40 % of flour and 60 % of sugar.

and is therefore called the invariant probability density. For instance, for Bernoulli's shift one has the following recursive rule:

$$\rho(x,t+1) = \frac{1}{2}\rho\left(\frac{x}{2},t\right) + \frac{1}{2}\rho\left(\frac{x}{2}+\frac{1}{2},t\right),$$

and the invariant probability density is constant in the interval $[0,1]$:

$$\lim_{t\to\infty} \rho(x,t) = \rho^{inv}(x) = 1\,.$$

It is rather natural, from an epistemic point of view, to accept the above probabilistic approach: the introduction of $\rho(\mathrm{x},0)$ can be seen as a necessity stemming from the human practical impossibility to determine the initial condition. For instance, in the case of Bernouilli's shift, knowing that the initial condition $x(0)$ is in interval $[x^*, x^*+\Delta]$, it is natural to assume that $\rho(x,0) = 1/\Delta$ for $x \in [x^*, x^*+\Delta]$, and 0 otherwise.

For t large enough (roughly $t > t_* \sim \ln(1/\Delta)$) one has the convergence of $\rho(x,t)$ toward the invariant probability distribution. Let us note that such a feature holds for any finite Δ, while t_* weakly depends on Δ, therefore we can say that $\rho^{inv}(x)$, as well the approach to the invariant probability density, sure are objective properties independent of the uncertainty Δ. Perhaps somebody could claim that, since it is necessary to have $\Delta \neq 0$, the above properties, although objective, still have an epistemic character. We do not insist further.

Figures 1a and b show, in a rather transparent way, how the approach to the $\rho^{inv}(x)$ is rather fast and basically independent on the $\rho(x,0)$.

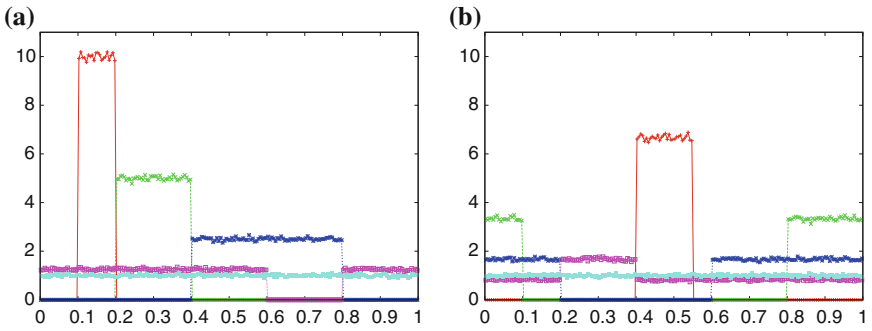


Fig. 1 Probability density $\rho(x,t)$ for Bernoulli's shift. The distribution is obtained generating at random 100,000 points, uniformly distributed in the interval $[0.1 : 0.2]$ (**a**) and $[0.4 : 0.55]$ (**b**), and then iterating the dynamics of each point for 14 time steps. The *red curves* correspond to the initial distribution, the *green*, the *blue*, and the *magenta curves* correspond to one, two and three time steps, respectively. As it is evident, at each time step the dynamics initially doubles the width of interval over which the points are distributed, until $\rho(x,t) \approx \rho^{inv}(x) = 1$. The light *blue curves* are obtained after 14 iterations: The probability density $\rho(x,t)$ for $t > 12$ is close to the invariant distribution $\rho^{inv}(x)$ independently of the initial $\rho(x,0)$

We saw how chaotic systems and, more precisely, those which are ergodic,[6] naturally lead to probabilistic descriptions, even in the presence of deterministic dynamics. In particular, ergodic theory justifies the frequentist interpretation of probability, according to which the probability of a given event is defined by its relative frequency. Therefore, assuming ergodicity, it is possible to obtain an empirical notion of probability which is an objective property of the trajectory (von Plato 1994). There is no universal agreement on this issue; for instance, Popper (2002) believed that probabilistic concepts are extraneous to a deterministic description of the world, while Einstein held the opposite view, as expressed in his letter to Popper:

I do not believe that you are right in your thesis that it is impossible to derive statistical conclusions from a deterministic theory. Only think of classical statistical mechanics (gas theory, or the theory of Brownian movement).

6 A Brief Digression on Models in Physics

At this point of the discussion, we wish to recall that our description of physical phenomena is necessarily based upon models,[7] that entail a schematization of a specified portion of the physical world. Take for instance the pendulum described by Eq. (1). The mathematical object introduced thereby relates to a physical pendulum under some specific assumptions. For instance, the string of length L, that connects the swinging body to a suspension point, is assumed to be *inextensible*, whereas any physical string is (to some extent) extensible. The model also assumes that gravity is spatially uniform and does not change with time, i.e., it can be described by a constant g. Equation (1) will therefore reasonably describe a physical pendulum only inasmuch as the variations of L and/or g are sufficiently small, so as to add only tiny corrections. Even more important, there is not in the physical world such an object as an isolated pendulum, whereas Eq. (1) totally ignores the physical world around the pendulum (the only ingredients being gravity, the string, and the suspension point). Galileo Galilei was well aware of this subtlety when comparing the prediction of our mathematical models with the physical phenomena they aim to

[6] A very broad definition of an ergodic system relies on the identification of time averages and averages computed with the invariant probability density (7). Said in other words, a system is ergodic if its trajectory in phase space, during its time evolution, visits (and densely explores) all the accessible regions of phase space, so that the time spent in each region is proportional to the invariant probability density assigned to that region. Therefore, if a system is ergodic, one can understand its statistical features looking at the time evolution for a sufficient long time; the conceptual and technical relevance of ergodicity is quite clear.

[7] There are several definitions of a *Model*, but to our purposes the following is a reasonable one: Given any system $\mathcal{S}$, by which we mean a set of objects connected by certain relations, the system $\mathcal{M}$ is said a model of $\mathcal{S}$ if a correspondence can be established between the elements (and the relations) of $\mathcal{M}$ and the elements (and the relations) of $\mathcal{S}$, by means of which the study of $\mathcal{S}$ is reduced to the study of $\mathcal{M}$, within certain limitations to be specified or determined.

describe[8] and called *accidents* (on this topic, see, e.g., Koertge 1977) all external influences apt to modify, often in an apparently unpredictable way, the behaviour of a (supposedly isolated) portion of the physical world. In the case of the Eq. (1), we are, e.g., neglecting the fact that a real pendulum swings in a viscous medium (the air), and also experiences some friction at the suspension point. These effects gradually alter the motion of the pendulum, which is no longer periodic and eventually stops. Equation (1) also neglects the fact that the Earth is not an inertial reference frame: it rotates around its axis and around the Sun. The first effect is far more important and gives rise to the gradual but sizable variation of the plane of oscillation (Foucault's pendulum). There are several external influences that may alter the motion of a pendulum. Some of them may be accounted for, at least to some extent, by simple modifications of Eq. (1). Other are rather complicated and are not easily accountable. Thus, Eq. (1) describes a pendulum only as far and as long as external influences do not alter significantly its motion. Said in other words, it describes a pendulum under *controlled conditions*.

7 The Old Dilemma Determinism/Stochasticity

The above premise underlines the crucial importance of the concept of *state of the system*, i.e., in mathematical terms, the variables which describe the phenomenon under investigation. The relevance of such an aspect is often underestimated; only in few situations, e.g., in mechanical systems, it is easy to identify the variables which describe the phenomenon. On the contrary, in a generic case, there are serious difficulties; we can say that often the main effort in building a theory of nontrivial phenomena concerns the identification of the appropriate variables. Such a difficulty is well known in statistical physics; it has been stressed, e.g., by Onsager and Machlup (1953) in their seminal work on fluctuations and irreversible processes, with the caveat:

how do you know you have taken enough variables, for it to be Markovian?

In a similar way, Ma (1985) notes that:

the hidden worry of thermodynamics is: we do not know how many coordinates or forces are necessary to completely specify an equilibrium state.

[8]Experiments are usually carried out under *controlled conditions*, meaning that every possible care is taken in order to exclude external influences and focus on specific aspects of the physical world. In his "Dialogues concerning two new sciences", Galilei (English translation 1914) describes the special care to be taken in order to keep the accidents under control: "… I have attempted in the following manner to assure myself that the acceleration actually experienced by falling bodies is that above described. A piece of wooden moulding or scantling, about 12 cubits long, half a cubit wide, and three finger-breadths thick, was taken; on its edge was cut a channel a little more than one finger in breadth; having made this groove *very straight, smooth, and polished*, and *having lined it with parchment*, also *as smooth and polished as possible*, we rolled along it a *hard, smooth, and very round* bronze ball …" (the italicized emphases are ours).

Unfortunately, we have no definite method for selecting the proper variables.

Takens (1981) showed that from the study of a time series $\{u_1, u_2, \ldots, u_m\}$, where u is an observable sampled at the discrete times $t_j = j\Delta t$ and $u_j = u(t_j)$, it is possible (if we know that the system is deterministic and is described by a finite dimensional vector) to determine the proper variable X. Unfortunately the method has rather severe limitations:

(a) It works only if we know a priori that the system is deterministic;
(b) The protocol fails if the dimension of the attractor[9] is large enough (say more than 5 or 6).

Therefore the method cannot be used, apart for special cases (with a small dimension), to build up a model from the data.

We already considered arguments, e.g., by van Kampen, which deny that determinism may be decided on the basis of observations. This conclusion is also reached from detailed analyses of sequences of data produced by the time evolutions of interest. In few words: the distinction between deterministic chaotic systems and genuine stochastic processes is possible if one is able to reach arbitrary precision on the state of the system.

Computing the so-called ϵ-entropy $h(\epsilon)$, at different resolution scales ϵ, at least in principle, one can distinguish potentially underlying deterministic dynamics from stochastic ones.

From a mathematical point of view the scenario is quite simple: for a deterministic chaotic system as $\epsilon \to 0$ one has $h(\epsilon) \to h_{KS} < \infty$, while for stochastic processes $h(\epsilon) \to \infty$.[10] On the other hand an arbitrary solution is not possible, therefore the analysis of temporal series can only be used, at best, to pragmatically classify the stochastic or chaotic character of the observed signal, within certain scales (Cencini et al. 2009; Franceschelli 2012). At first, this could be disturbing: not even the most sophisticated time-series analysis that we could perform reveals the "true nature" of the system under investigation, the reason simply being the unavoidable finiteness of the resolution we can achieve.

On the other hand, one may be satisfied with a non-metaphysical point of view, in which the true nature of the object under investigation is not at stake. The advantage is that one may choose whatever model is more appropriate or convenient to describe the phenomenon of interest, especially considering the fact that, in practice, one observes (and wishes to account for) only a limited set of coarse-grained properties.

In light of our arguments, it seems fair to claim that the vexed question of whether the laws of physics are deterministic or probabilistic has, and will have, no

[9]The attractor of a dynamical system is a manifold in phase space toward which the system tends to evolve, regardless of the initial conditions. Once close enough to the attractor, the trajectory remains close to it even in the presence of small perturbations.

[10]Typically $h(\epsilon) \sim \epsilon^{-\alpha}$ where the value of α depends on the process under investigation (Cencini et al. 2009).

definitive answer. On the sole basis of empirical observations, it does not appear possible to decide between these two contrasting arguments:

(i) Laws governing the Universe are inherently random, and the determinism that is believed to be observed is in fact a result of the probabilistic nature implied by the large number of degrees of freedom;
(ii) The fundamental laws are deterministic, and seemingly random phenomena appear so due to deterministic chaos.

Basically these two positions can be viewed as a reformulation of the endless debate on quantum mechanics: thesis (i) expresses the inherent indeterminacy claimed by the Copenhagen school, whereas thesis (ii) illustrates the hidden determinism advocated by Einstein (Pais 2005).

References

Atmanspacher, H.: Determinism is ontic, determinability is epistemic'. In: Atmanspacher, H., Bishop, R. (eds.) Between Chance and Choice. Imprint Academic, Thorverton (2002)
Boyd, R.: Determinism, laws and predictability in principle. Phylosophy Sci. **39**, 43 (1972)
Campbell, L., Garnett, W.: The Life of James Clerk Maxwell. MacMillan and Co., London (1882)
Cencini, M., Cecconi, F., Vulpiani, A.: Chaos: From Simple Models to Complex Systems. World Scientific, Singapore (2009)
Chibbaro, S., Rondoni, L., Vulpiani, A.: Reductionism, Emergence and Levels of Reality. Springer-Verlag, Berlin (2014)
Dyson, F.: Birds and frogs. Not. AMS **56**, 212 (2009)
Franceschelli, S.: Some remarks on the compatibility between determinism and unpredictability. Prog. Biophys. Mol. Biol. **110**, 61 (2012)
Galilei, G.: Dialogues Concerning Two New Sciences. MacMillan, New York (1914)
Koertge, N.: Galileo and the problem of accidents. J. Hist Ideas **38**, 389 (1977)
Lorenz, E.N.: Deterministic nonperiodic flow. J. Atmos. Sci. **20**, 130 (1963)
Ma, S.K.: Statistical Mechanics. World Scientific, Singapore (1985)
Onsager, L., Machlup, S.: Fluctuations and irreversible processes. Phys. Rev. **91**, 1505 (1953)
Pais, A.: Subtle is the Lord: The science and the life of Albert Einstein. Oxford University Press, Oxford (2005)
Poincaré, H.: Les méthodes nouvelles de la mécanique céleste. Gauthier-Villars, Paris (1892)
Popper, K.R.: The open universe: An argument for indeterminism. From the Postscript to the Logic of Scientific Discovery. Routledge, London (1992)
Popper, K.R.: The Logic of Scientific Discovery. Routledge, London (2002)
Prigogine, I., Stengers, I.: Les Lois du Chaos. Flammarion, Paris (1994)
Primas, H.: Hidden determinism, probability, and times arrow. In: Bishop, R., Atmanspacher, H. (eds.) Between Chance and Choice, p. 89. Imprint Academic, Exeter (2002)
Shannon, C.E.: A note on the concept of entropy. Bell Syst. Tech. J. **27**, 379 (1948)
Stone, M.A.: Chaos, prediction and Laplacean determinism. Am. Phylos. Q. **26**, 123 (1989)
Takens, F.: Detecting strange attractors in turbulence. In: Rand, D., Young, L.S. (eds.) Dynamical Systems and Turbulence, Lecture Notes in Mathematics **898.** Springer-Verlag, New York (1981)
van Kampen, N.G.: Determinism and predictability. Synthese **89**, 273 (1991)
von Plato, J.: Creating Modern Probability. Cambridge University Press, Cambridge (1994)
Vauclair, S.: Elémentes de Physique Statistique. Interéditions, Paris (1993)

Ways of Advancing Knowledge. A Lesson from Knot Theory and Topology

Emiliano Ippoliti

Abstract The examination of the construction of several approaches put forward to solve problems in topology and knot theory will enable us to shed light on the rational ways of advancing knowledge. In particular I will consider two problems: the classification of knots and the classification of 3-manifolds. The first attempts to tell mathematical knots apart, searching for a complete invariant for them. In particular I will examine the approaches based respectively on colors, graphs, numbers, and braids, and the heuristic moves employed in them. The second attempts to tell 3-manifolds apart, again searching for a complete invariant for them. I will focus on a specific solution to it, namely the algebraic approach and the construction of the fundamental group, and the heuristic moves used in it. This examination will lead us to specify some key features of the ampliation of knowledge, such as the role of representation, theorem-proving and analogy, and will clear up some aspects of the very nature of mathematical objects.

1 Introduction

The classification of knots and the classification of manifolds are long-standing problems in knot theory and topology (see e.g. Adams 2004) that offer an interesting chance to study the rational ways by which knowledge is advanced and to contribute to the ongoing study of these ways.[1] They are a sort of laboratory of mathematical activity, where we can see how problems are found, posed, and

[1]See in particular: Polya (1954), (Hanson 1958), (Lakatos 1976), (Laudan 1977), (Simon 1977), (Nickles 1980a, b), (Simon et al. 1987), (Gillies 1995), (Grosholz and Breger 2000), (Abbott 2004), (Darden 2006), (Weisberg 2006), (Magnani 2001), (Nickles and Meheus 2009), (Magnani et al. 2010), (Cellucci 2013), (Ippoliti 2014).

E. Ippoliti (✉)
Sapienza University of Rome, Rome, Italy
e-mail: emi.ippoliti@gmail.com

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_9

solved, and how the tools to do it are created and progressively refined. Moreover, in this laboratory we can examine in vivo (or at least in vitro) how solutions and hypotheses are generated, and observe the interplay between mathematics and natural sciences (see also Ippoliti 2008, 2011, 2013). In particular topology and knot theory offer tools for tackling those problems where form and function are strictly related. Most of these problems originate from physics and natural science (i.e. chemistry, biology) and mathematical entities are developed in order to deal with them.

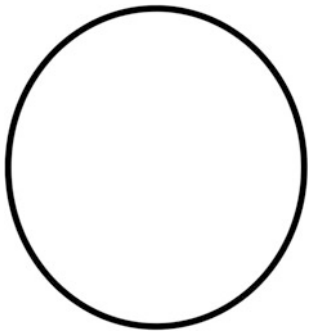
As I will show, the attempts to solve the problems of the classification of knots and of the classification of manifolds tell stories of success and failure, or better of partial success and failure, but we can learn a lot also from failures, or partial failures. It is just when an attempt of solution does not succeed that it displays its own inner dynamics: we can reconstruct the rational and inferential steps of this process and better see its interpretations and manipulations of objects.

2 Invariants in Knot Theory

The problem of classification of knots originated in nineteenth century theoretical physics, as Hermann Von Helmholtz, William Thomson (Lord Kelvin), Maxwell, and Peter Guthrie Tait tried to develop a theory about atoms. The climax of this effort was Lord Kelvin's vortex theory (Thomson 1869). As well known, it conjectures that atoms are knotted tubes of ether, and has been received for about two decades. In particular, the vortex theory draws on the hypothesis that the stability of matter and the variety of chemical elements derive respectively from the topological stability and the variety of the underlying knots. Accordingly, this theory needed a mathematical foundation, that is a study and classification of knots capable of telling us if two knots are different or not. In this way the problem passed from physics to mathematics: Tait (1900) put forward an extensive study and tabulation of knots in the attempt to determine when two knots are isotopically different. In effect, if vortex's theory turns out to be the right basis for the classification of the chemical elements, then a knot table is needed to found a periodic table of elements. Unfortunately Kelvin's theory did not succeed and was abandoned. But the mathematical problem survived and continued to draw the interest of mathematical community.

A mathematical knot is a simple, closed, non-self-intersecting curve in R^3 and the mathematical study focuses on its behavior under a specific kind of deformation—ambient isotopy: you can distort a knot (stretch, pull, etc.) without breaking it. You cannot cut and glue it. The classification of knots aims at determining if two knots are equivalent under isotopy, in order to distinguish different types of knots. The solution of this problem requires finding a way to tell non-equivalent knots apart. The first step to do this it to find an existence proof of knots, by specifying a procedure capable of telling us when a given knot is equivalent to the unknot or not (see Fig. 1)—that is, if it is possible or not to reduced it to a circle under isotopy.

Fig. 1 The unknot



Classifying knots up to equivalence turned out to be a very hard problem. The basic strategy to tackle it was the search for invariants, by setting up a procedure capable of assigning the same 'quantity' to equivalent knots. More specifically the goal was to find ways of defining and associating a quantity to knots such that it does not change under isotopic deformations if the knots are the same. In order to find a complete invariant, several approaches have ben put forward. I will examine four of them: the ones based respectively on colors, numbers, graphs, and braids. I will show that these approaches are ways of interpreting and manipulating knots, and how different approaches reveal both specific features of knots and specific features of ways for ampliating knowledge.

Tellingly, even though knots are 3d objects, they are opaque to most of the tools developed for investigating 3d objects. This fact motivated the attempt to employ tools from other fields in order to deal with the problem. But in order to that, a change of representation was needed. More specifically, the construction of a suitable 2d representation of knots has permitted to use tools and results for 2d objects to shed light on knots. In this sense, it is worth noting that the applicability of tools and knowledge from another domains motivates the choice of a particular representation of the mathematical object. The representation of a knot in a 2d format is a crucial and not so easy task to perform. It requires a set of devices, interpretations, and changes of a knot that can generate new problems—as it did. It is not a neutral move and shows how a key tool in problem-solving is changing representation. In this case, this step is accomplished with the construction of a knot diagram, i.e. the representation of a knot by means of a suitable projection into a plane (see Fig. 2). Roughly, a knot diagram is the shadow of the knot plus the crossing information. More precisely it is the image of a function from the R^3 to the plane, taking the triple (x, y, z) to the pair (x, y), and which meets certain conditions. In effect there are several possible ways of constructing projections of a knot, so we have to make choices. Some projections are better than others, for some of them involve too much loss of information and it would hard to reconstruct the 3d knot from these pictures. For instance all the projections in Fig. 2b (called irregular) would make the study of knots very hard. On the contrary, we simplify the investigation by focusing on a specific knot projection (called regular, see Fig. 2a)

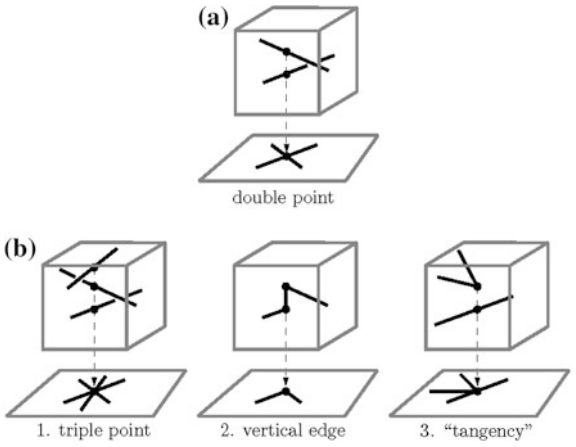


Fig. 2 **a** Regular projection of a knot, **b** irregular projections of a knot

where three points in the knot cannot project to the same point, and no vertex projects to the same point as any other point on the knot. A regular projection of a knot is called knot diagram (see Fig. 3).

This simple operation introduces new information in the problem (i.e. not contained in its 3d formulation): we define crossings and arcs in a knot diagram. A crossing is a place where the projections of the knot curve crosses–going over or under–itself. An arc is a piece of the curve going between two undercrossings—overcrossings are allowed along the way.

The study of diagrams, rather than 3d curves, makes it possible the application of new tools in problem-solving, enhancing our ability to tackle it. But it also generates new problems, since part of information of the 3d object is neglected. An example of new problem is the fact that the same knot can be represented by several diagrams (depending on the projection point, see Fig. 4) and, hence, it can be difficult to establish if two or more diagrams represent the same knot.

Since there are an infinite number of moves that can be performed on a particular diagram, is it possible to establish if two given diagrams represent the same knot? This point is crucial: it raises the question of whether a particular projection determines what we can infer. If we cannot answer this question, we cannot rely on the study of knot diagrams to shed light on knots and their properties. The

Fig. 3 A knot diagram

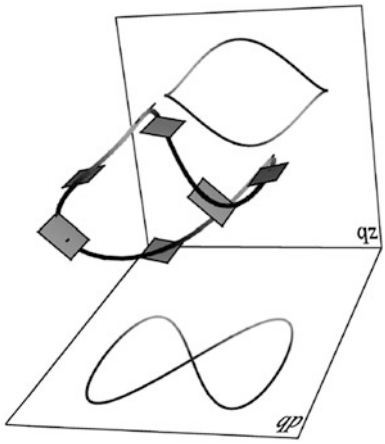


Fig. 4 Two 2d projections of the same knot

Reidemeister theorem (Reidemeister 1927) answers the question: the results are not diagram-sensitive. We need only three moves to go from one projection of a knot to any other projection of the same knot, regardless of how complicated is the sequence of these three moves. More formally the theorem tell us that two diagrams are ambient isotopic if and only if they are connected by a finite sequence of the three Reidemeister moves (see Fig. 5).

The Reidemeister theorem holds also for oriented diagrams (and links), so it accounts for all the possible coherent orientations of diagrams. But even if the Reidemeister theorem provides us with a powerful tool to deal with knots by mean of their diagrams, it a has a limit: with these three moves you can tell whether two knots are the same, but you cannot tell them apart, if they are different.

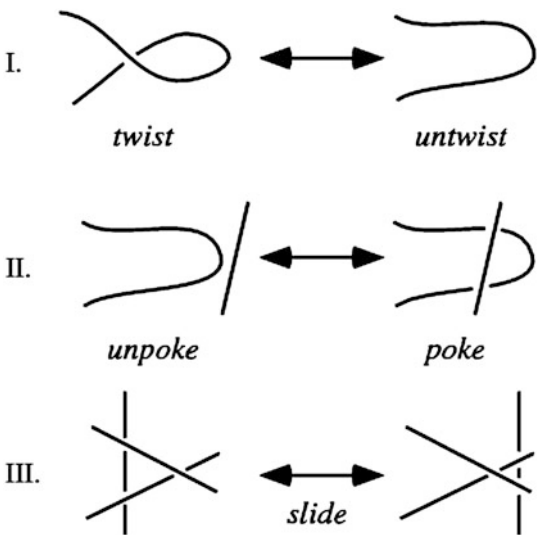


Fig. 5 The Reidemeister moves

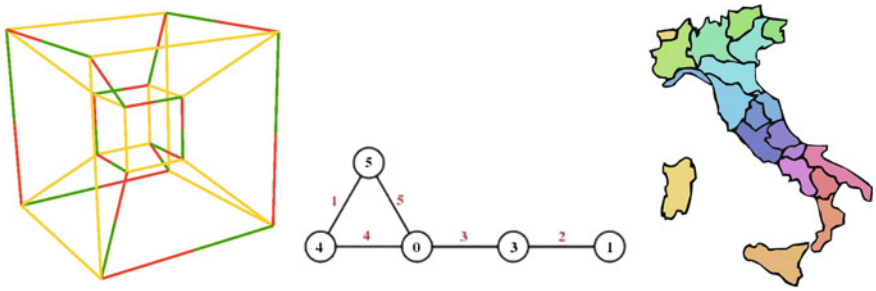


Fig. 6 Three ways of labeling

2.1 Coloring Knots: Heuristic Construction of New Representations

Coloring is a technique developed in '800 for dealing with problems in discrete mathematics (e.g. 4-colors or 5-colors theorems). Coloring is a way of assigning distinct labels (a color, a number, a letter, etc.) to each component of a discrete object (see Fig. 6), such as a plane or a graph. It makes possible to pose and solve several problems involving topological properties. In knot theory, coloring has been introduced by Crowell and Fox (1963[2]) in order to tell knots apart, and since then it has extensively used (see Kauffman 1991; Montesinos 1985).

From a heuristic point of view, the salient aspect of the application of coloring to knot theory is the fact that, *strictu sensu*, there are no components for knots. A knot is a single strand in 3d space, and it has no crossings and strands: consequently, it cannot be discretized. Thus no labels can be identified, as there is only one item—the single string in 3d. Accordingly, coloring is a non-sense for 3d knots: it cannot be even defined for them. On the other hand, a 2d object, like a knot diagram, can be discretized and, hence, colored. It is just after this change of representation that coloring can be applied to the study of knots (see Fig. 7). For instance, we can employ two well-known techniques like 3-colorability or 5-colorability. Let us focus on 3-colorability. A diagram D is tricolored if:

1. every arc is colored, e.g., r (red), b (blue) or y (yellow);
2. at any given crossing appear either all three colors or only one color.

Now, 3-coloring turns out to be an invariant, event tough a basic one. It allow us to tell several knots apart. For instance, 3-coloring is the simplest invariant that distinguishes the trefoil knot and the trivial knot. Fox shows that every 3-colorable knot is nontrivial. More precisely he proves that if a given diagram can be 3-colored, then it expresses a knot different from the unknot, which cannot be 3-colored since only once color can be assigned to its one arc. In other words, we can tell apart any 3-colorable from non 3-colorable knot. Accordingly, since we can

[2]See in particular Chap. 6, exercises 6–7.

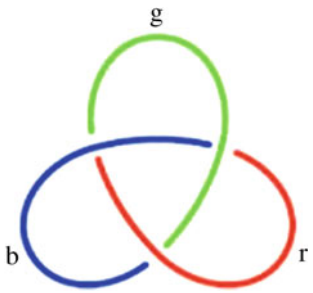


Fig. 7 Tricolored knot diagram

certainly 3-color some diagrams, then we have just proven the existence of non-trivial knots. Moreover it distinguishes the trefoil knot and its mirror images, or the trefoil knot and the unknot. On the other hand, 5-colorability tells apart the figure eight knot and its mirror image. But coloring does not provide a complete invariant. For instance 3-colorability cannot distinguish the figure eight knot and its mirror images, or figure-eight knot and the unknot. And 5-colorability cannot tell apart the trefoil and its mirror image.

So even if colorability cannot be even defined for knots in 3D, it ends up revealing properties of knots, and not simply of their diagrams. This is not obvious, since a new representation adds and cut offs features of the original entities that affect the results obtained by the use of other tools.

2.2 Graphing Knots: The Heuristic Change of Representation

There are a number of ways of manipulating and interpreting a mathematical object to make it treatable by specific tools in order to advance hypotheses for solving a given problem. This operation does not come cheap, not to say free. It has a specific cost: these interpretations often requires a change of the representation of the object and this, in turn, implies a choice about which characteristics to highlights and which one to ignore, which ones are to be considered relevant and which ones negligible. It goes without saying that this is a tentative move, which can be justified only on heuristic basis. Coloring ignores a set of 3-dimensional properties of knots, using a specific change in representation and it is designed to focus on 2d discrete characteristics, so to use approaches developed for discrete mathematics in the search for a classification of knots.

Another way of conceptualizing knots by means of diagrams is by using graph. A *graph* is finite set $V(G)$ of vertices and a finite set of edges $E(G)$—see Fig. 8. Any edge is paired with vertices that are not necessarily distinct and are called endpoints of the edge. Moreover, a graph G can have multiple edges and loops.

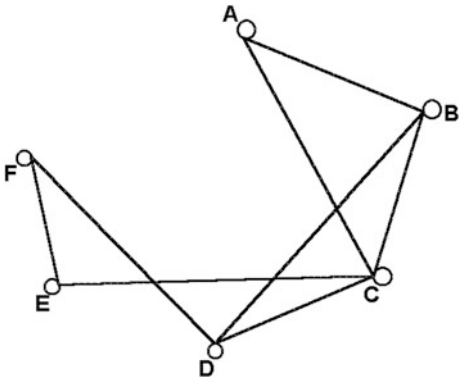


Fig. 8 Example of graph

The conceptualization of knots by means of graphs is based on an analogy, that is the fact that graphs and knots are both 2d objects, namely closed plane curves, and as such they can be interpreted, once suitably manipulated, as ways of separating a plane in regions. This requires manipulating both graphs and knots in a very specific way in order to make them similar under certain respects and employ graph theory in the study of knots.

Tellingly, Tait himself was the first to read knots as planar graphs. More precisely, he manipulated a knot diagram in the following way (Tait 1887, 292–94): he colored the regions of the diagram alternately white and black (following Listing 1847) and constructed the graph by placing a vertex inside each white region and then connecting the vertices by edges passing through the crossing points of the diagram (see Fig. 9).

Now we can complete this transformation by expressing the crossing of a knot simply introducing ± values at each crossing, obtaining a *signed planar graph*. At the end of this manipulation, i.e. a change or adaptation to a specific purpose, we have a new object, namely the *graph of a knot* (see Fig. 10). The bottom line: we have just generated a new viewpoint on knots and we can investigate them with the tools provided by graph theory.

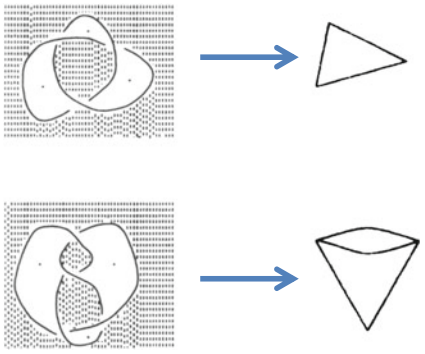
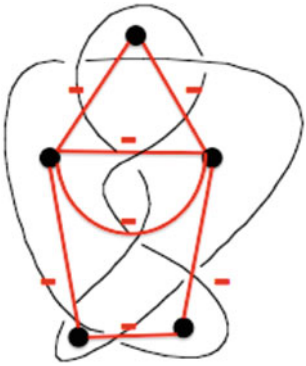


Fig. 9 Obtaining a graph from a knot

Fig. 10 Graph of a knot



This heuristic manipulation offers several contributions to the study of knots. The three main contributions of graph theory to the study of knots are (e.g. Yajima and Kinoshita 1957; Kinoshita and Terasaka 1957):

1. The deformation of knots, i.e. the equivalence of knots, can be explained schematically by the graphical representation of knots.
2. The graphical representation of knots contributes to the study of construction of knots.
3. Graphs knots provide a necessary and sufficient condition for amphibious knots.

For instance in the latter, graph theory allows us to prove a theorem stating that if the knot graphs $g(\pi)$ and $g^*(\pi)$ are of same type and have opposite signs, then the original knot K of π is amphibious. That is, K is equivalent to its mirror image (Reidemeister 1932; Schubert 1956).

2.3 Arithmetizing Knots: Cutting the Problem Space

Another way of attacking the problem of classification of knots is to investigate them in terms of *composition*, i.e. a number-theoretical operation that classifies objects by ordering them from the simpler to the more complicated by means of a function of composition (and decomposition).

An efficient way of doing this is factorization. It offers a straight heuristic advantage: it does not require to classify all knots, but rather only those that cannot be made up of smaller "pieces" or knots—the prime factors. The heuristic move of interpreting knots as numbers, that is numbering knots in a way the preserve the decomposition into suitably defined *prime factors*, is the seminal idea of the work of Alexander and Briggs (1926–27). With this new representation, the search for a solution to the problem of classification of knot is reduced to the search for a way of defining and associating a prime number to a prime knot and a composite number to a composite knot, so that the prime factors of the number are the prime factors of the knot. In effect, if the approach of finding a unique factorization of knots into

prime components succeeds, we would get an important result: a drastic reduction in the search for invariant for knots. First, instead of having to look for invariants of all knots, we could focus only on invariants of prime components—and how those invariants behave under the connected sum. Second, we could focus on the smaller components that make up knots. This would produce a considerable simplification in the study of knots, as we could totally ignore the order by which a knot is broken into prime components—since the resulting factors will be the same.

Number-theory offers just this way of classifying objects, and this motivates the search for a way of treating knots as numbers. In other words we are searching for a way of transforming the target (knots) so to make them treatable by our source (numbers), in order to mimic into knots as many relevant properties of number factorization as possible. The problem is whether there is a way to manipulate knots that permits this number-theoretical conceptualization, by constructing and defining prime knots, composite knots, and composition between knots. This is a very strong heuristic move: in essence we are trying to build new entities, relations, and operators on knots.

The idea of a number-theoretical approach to knots goes back to Gauss (1798), who used the analogy between primes and knots, and it was put forward by Schubert (1949) and Mazur (1973). In principle, there are several possible ways of constructing a number-theoretical version of a knot. To this end, it is worth recalling that the crucial property of a knot is not its manifold structure (under which every knot is equivalent to a circle) but rather its embedding into the ambient space. So any attempt of defining the operation of composition of knots (the connected sum) requires an appropriate definition that produces a well-defined embedding. A way of doing this is to cut each knot at any point and join the boundaries of the cut, keeping orientations consistent (see Fig. 11).

This definition of composition offers two advantages: it is independent of the location of the cut and does not generate new crossings. Thus we are now in a position to pose and solve problem in knot theory by analogy with number-theoretical well-known properties of composition. In particular the analogy allows us to explore the features of composition like commutativity, associativity, subtraction, inversion, and factorization. I denote composition with the symbol #. Schubert (1949) proved that commutativity holds for knots (see Fig. 12)—namely given two knots a and b, $a\#b = b\#a$.

Moreover, he showed that also associativity holds for knots: given the knots a, b, and c, $(a\#b)\#c = a\#(b\#c)$. On the other hand, there is no way to define the

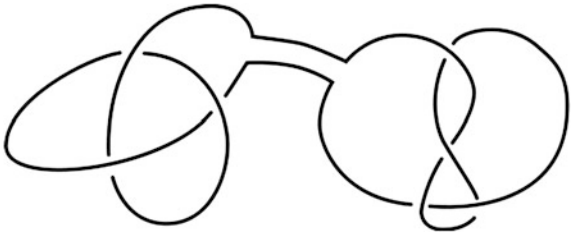


Fig. 11 Composing knots

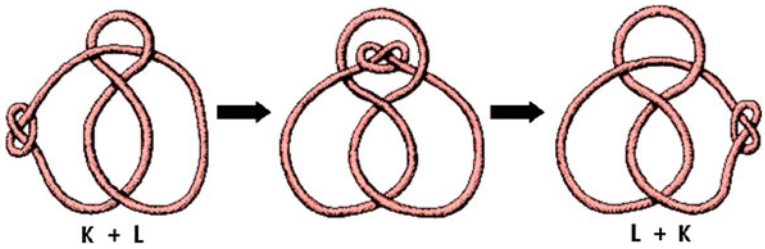


Fig. 12 Commutativity for knots

subtraction of knots. The arithmetical analogy fails here. Tellingly, just like natural numbers, where for every natural number $n > 1$ there is no natural number m such that $n\#m = 1$, it is not possible to find the inverse knot, that is given a knot to find another knot that, composed with it, gives the unknot. More precisely there is no *general* way of finding a knot that composed to another ones cancel them out (gives the unknot)—even if you can find it for specific instances.

Once the features of composition have been explored and clarified for knots, it is possible to employ one of its main tools—i.e. factoring. Since this requires building blocks, the primes, the first step of our inquiry is to find a way of constructing the counter-part of *primes* for knots, which requires identifying them (show their existence) and, in case, their features. Just like prime numbers are the product of smaller numbers, a prime knot can be conceived as one that is not the sum of simpler knots. First, it turns out that it is possible to prove the existence of prime knots. It can be done in several ways—for instance using knot genus. So, a knot is called prime if it cannot be represented as a connected sum of two knots that are both knotted. Second, factoring turns out to be a more complex operation than connected sum to define and perform. The standard way of factoring a knot, say K, is by using a sphere, S, that cuts K at two points and then separates it into two components K_1 and K_2—joining the two loose ends of each knot with some path in the sphere (see Fig. 13). The result of this operation is problematic, since it depends on the sphere chosen to factor the knot.

It is worth noting that a knot can be decomposed recursively, by iteratively factoring it into smaller and smaller components ending up with prime components. Factoring knots turns out not only to be possible, but unique up to the order. Schubert's theorem (Schubert 1956) just states that every knot can be uniquely

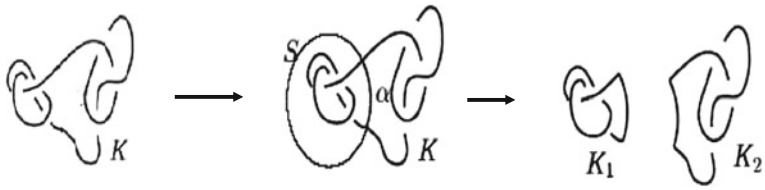


Fig. 13 Decomposing a knot in factors

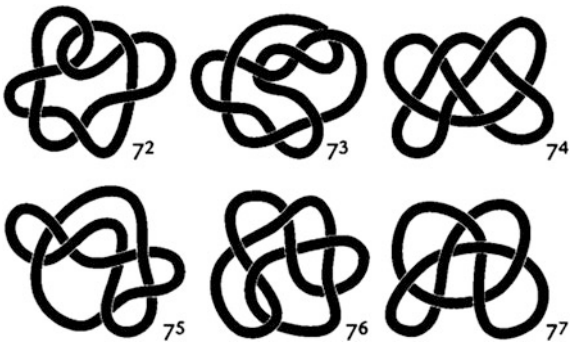


Fig. 14 Ordering knots by minimal crossings

decomposed in a finite number of prime knots. In this way we have produced an *arithmetic* of knots, which shows some essential features of knots. Thus this approach offers a genuine advancement in the search for a solution of knot classification by drastically reducing the space of research for invariant for knots. Unfortunately, it does not solve completely the problem. Natural numbers have a total order that does not exist for knots—or at least has not been discovered. Of course, few attempts of fixing this point and ordering knots have been put forward. The received one is ordering knots by the number of minimal crossings of their diagrams. But this order is not linear, and does not allows us to order knots like the ones in Fig. 14—that is answering the question about which one is the smallest.

3 Notations as Heuristics: The Conway's Case

The seemingly primitive and neutral act of notating turns out to be a way of ampliating knowledge. More precisely, a notation is the endpoint of a specific conceptualization and a heuristic tool employed in problem-solving. Knot theory provides a particularly strong example in this sense, namely the Conway's notation for knots (Conway 1967), which offers an explicit example of the heuristic role of notation. Conway himself states that he built "a notation in terms of which it has been found possible to list (by hands) all knots of 11 crossing or less and […] some properties of their algebraic invariants whose discovery was a consequence of the notation" (Ibid., 329).

In essence, Conway notation is a way of representing a knot by means of a sequence of integer numbers that express a selection of its properties. The construction of this notation provides us a magnifying glass over the rational steps put forward to advance knowledge. It uses two heuristics. First, the use of knot diagrams. Second, the reduction of a global problem to a local one. In latter case Conway focuses on the properties of portions of knot-diagram, called *tangles*. A tangle is a part of a knot projection—namely a circled region in the projection plane such that four strands exit the circled region (see Fig. 15). Then he shows that a whole class of

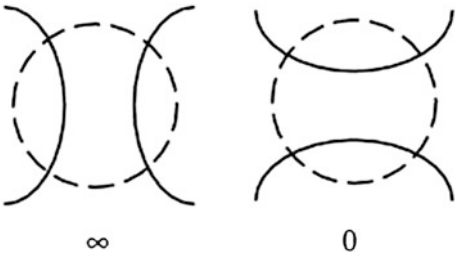


Fig. 15 Infinity-tangle and zero-tangle

knots can be represented by tangles, by performing simple changes on tangles, ending up with the Conway theorem. The definition and construction of *tangles* is not arbitrary: in particular the fixed number of strands exiting the circle draws on the idea of combining tangles and performing arithmetical operations. In effect, Conway shows how to represents tangles with numbers by using a simple algorithm.

So, Conway's notation uses three basic tangles in order to generate more complicated ones: the infinity-tangle, the zero-tangle, and the n-tangle, where n is an integer. For instance, a 3-tangle is a tangle with 3 twists in it (see Fig. 16). Then the notation introduces a distinction between $-n$ tangles and $+n$ tangle by means of overcrossings. More precisely, if a overcrossing has a positive slope, the tangle has a positive value; if it has a negative slope, the tangle has a negative value. On this basis, Conway produces *rational tangles* by preforming a sequence (an algorithm) of simple operations on tangles—such as rotation, twist and reflection. This procedure associates a sequence of numbers to tangles. For example, you get the 2 3 tangle by starting from a 2 tangle, rotating it and joining the outcome of rotation with its mirror image. The numbers keep track of these properties of the rational tangles and can be used to classify them (tell them apart). Conway shows how to represent any knot by means of a tangle: given any tangle you can transform it in a knot by joining its two northern and two southern strands together. This seemingly simple construction is the heuristic move that allows us to extend to knots results and tools for tangles and, in particular, offer a way to classify them.

Thus the question that naturally arises here is if different notations do always expresses different tangles. The answer, unfortunately, is no. As a matter of fact, it is possible to show, e.g., that the 2 2 1 tangle and the −2 2 1 tangle are equivalent (under Reideimester moves). The bottom line: notation is not a complete invariant for tangle and, accordingly, for knots. Nevertheless, this notation is the starting point of another notation that allowed Conway to achieve his most important contribution to knot theory. This new notation employs continued fractions and

Fig. 16 3-tangle

ends up with the Conway's theorem, which states that two rational tangles are equivalent if and only if their continued fraction values are the same. For instance −2 3 2 tangle and 3 −2 3 tangle are equivalent. Remarkably, the continued fraction value embodies the shape of the tangle itself—and of course knots formed from these tangles will also be identical.

The continued fraction representation of tangles uses the indexes of Conway's notation, i.e. the sequence of numbers. For instance, let us take the 5 1 4 tangle and the 2 −2 2 −2 2 4 tangle. Apparently the numbers tell us that they are distinct tangles. The continued fractions representations of these two tangles are respectively:

$$5+\frac{1}{1+\frac{1}{4}}, \quad \text{and} \quad 2+\frac{1}{-2+\frac{1}{2+\frac{1}{-2+\frac{1}{2+\frac{1}{4}}}}}$$

A rapid calculation shows that their continued fractions value is the same (29/6). Hence, in virtue of the Conway theorem we know that they expresse that same tangles and so are the knots obtained from them. This means that we can perform a sequence of Reidemeister moves that allows us to transform one tangle in the other, and vice versa.

Moreover, using Conway notation it is possible to prove that all the knots obtained from rational tangles are alternating knots. So given an alternating knot, we can determine most of its properties by means of rational tangles theory. We can show that also the definitions, not only the notations, can play a similar heuristic role (see e.g. Kauffman 1987). A nice example is the X polynomial for knots (Kauffman 1991, 33), but I won't treat it here. I merely point out that this example displays that also definitions are laden with ampliative reasoning and that they are not mere stipulations, but end-points of conceptualizations and interpretations of given objects.

4 Braiding Knots: Theorem-Proving and Heuristics

An essential passage in constructing Conway's notation is a *representational* one, a result stating that a given object with certain properties is isomorphic to another object—namely that knots and tangles are equivalent under certain respects. This passage is crucial for ampliation of knowledge in general: it formally enables us to employ results and tools produced for the one object in the study of the other. When this assimilation is established explicitly, a set of concepts, operations and properties can be transferred from the one to the other, reducing and shaping the space of research for solutions of a problem.

This process displays the heuristic role of theorem-proving: the continual search of mathematicians for a way of representing a given structure by means of another is a tool for producing new knowledge by connecting two unrelated objects. The connection is new—not the objects. This also explains why certain objects are

manipulated in a way rather than in another way: one way is approachable by pieces of known mathematics, whilst the other not. Essentially theorem-proving is a method of reduction in two senses. First, a reduction of a selection of certain properties of one object to (a selection of properties of) another objects. Second, the reduction of the space of research for a solution.

Knot theory offers a strong example of the heuristic role of theorem-proving: the Alexander's theorem (Alexander 1923). This representation theorem simply states that every knot has a closed braid presentation—that is every knot is isotopic to a closed braid. I won't give the detail of the theorem, but I will examine two important consequences of it:

1. it formalizes a new relation between two entities—knot diagram and the closure of braids. So if you have full understanding of braids, it turns out that you will have a good understanding of knots (and links).
2. Since braids are conceptualized in algebraic terms (i.e. group theory), it turns out that algebra con be used to study knots too.

A braid is a set of n strands attached between two horizontal bars. The strands must always go down from the top horizontal bar to the bottom horizontal bar and can cross each other (see Fig. 17).

In essence, a braid is a two-dimensional object expressing information about the relative position of the strands and their overcrossings—just like knots under this respect. Braids have been approached from an algebraic viewpoint: the *braid group* (Artin 1947) gives an deep understanding of braids in pure algebraic terms.

Since braids and knot diagrams are two ways of expressing similar information, the attempt to use the formers to understand the latters has been extensively put forward. But in order to establish this connection it is necessary to find a way of assimilating knots to braids, and to understand to what extent this is possible. This step has been accomplished with the operation of closure of a braid, a move that evolves up to the proof of Alexander theorem. The closure of braids b, denoted $CL(b)$, is simply the connection of the endpoints of the strands in a standard way, i.e. without introducing further intersections between the strings (see Fig. 18). It is possible to show that equivalent braids express equivalent knots, but a number of braids may express the same knot.

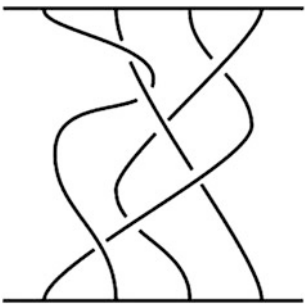


Fig. 17 A braid

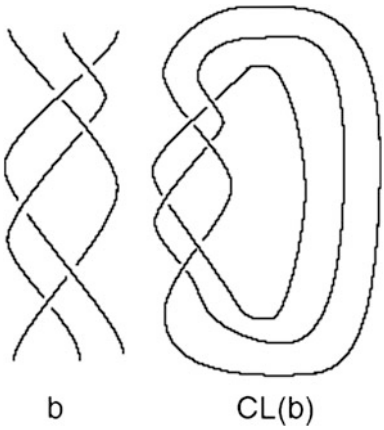


Fig. 18 Closure of a braid *b*

The assimilation of knots and braids produces a real advancement of knowledge in knot theory: not only it opens the door to the Alexander's theorem, but it also lays the foundation for the Brunn's theorem (Brunn 1897), stating that every knot has a projection with only one multiple point.

5 Invariants and Manifolds: The Fundamental Group

Because the classification of 3-manifolds requires extra levels of detail, it makes it possible to zoom in even closer on the rational ways of ampliating knowledge. This problem is hard, as traditional tools and concepts such as genus, orientability and boundary components cannot offer a complete description of 3-manifold. A key idea to classify them, again, is the search for invariants. But traditional invariants, like Euler characteristic, are useless for 3-manifolds since their values all became zero. Thus, a new approach is needed to tackle the problem and it was developed by employing algebra. This move draws on a simple heuristics: an invariant is something that does not change under certain operations—in this sense it is a structure. And algebra is just the study of structures (i.e. something preserved under specific functions or operations). So if we find a way of associating algebraic structures over the topological entities, we could be in a position to attack and solve the problem of classifying 3-manifolds: if a structure can be uniquely associated to a topological space and preserved, then it can tell manifolds apart. The problem, naturally, is which topological structure, if any, can be associated to topological entities as an invariant and, above all, how to do it. On second thought, this approach is not new: "elementary analytic geometry provides a good example of the application of formal algebraic techniques to the study of geometrical concepts" (Crowell and Fox 1963, 13). Poincaré put forward just this approach (Poincaré 1895), ending up with the construction of the fundamental group (or the first homotopy group).

The big question hence is how to build a bridge from algebra to topology. There are a number of ways of associating algebraic structures to topological spaces, many possible interpretations, and the data not only do not determine, but also do not suggest the more appropriate algebraic structure to build it. In order to do this, we have to manipulate a topological space, to introduce specific definitions, notations, operations and relations that are not contained in it. We have to conceptualize it in a new way, entering new and uncertain lands—the lands of heuristic reasoning.

The fundamental group is the endpoint of a step-by-step construction that gradually introduces new information into topological spaces (new entities, definitions, operations and relationships) by means of a rational manipulation of it, and using ampliative inferences. As the name suggests, the algebraic structure associated to a topological space at the end of this construction is a group—i.e. a set of elements with a composition map and three properties: associativity, an identity element, and an inverse element. It provides an algebraic structure that can 'measure' shapes by a calculation that maps numbers into topological spaces.

Before examining the rational and inferential content of this construction, and why just a group and not another algebraic structures is chosen, it is important to note that this process is not the mere search for an isomorphism, but a new interpretation and conceptualization of topological space.

To build a bridge between algebra and topology we have to answer at least two preliminary questions:

1. how to define the topological counterpart of the set of elements of an algebraic structure;
2. how to define the topological analogous of operations in algebraic structures.

First, since the Euler characteristic does not help us in differentiating 3-manifolds, we need a new kind of equivalence. This is *homotopy*, which can be defined as continuous deformation: in addition to pulling and stretching, it is possible to compress, a operation that changes dimension. Under homotopy, a disc, a ball and a point are all equivalent.

Second, we have to manipulate a topological space X in a very specific way in order to let it act like an algebraic structure. It is possible to construct several algebraic representations of topological entities. For instance, a first attempt could be made by looking at the set of all paths of a topological space between two given points, and then by defining a composition—e.g. their product (see Fig. 19). The paths would keep track of the information about the shape of the space, and their product would model their behaviour under homotopy. Unfortunately, the "meager algebraic structure of the set of all paths of a topological space with respect to the product" (Crowell and Fox 1963, 15) is far from offering something useful for the solution of our problem: it does not tell apart even simple distinct surfaces.

We need a better interpretation and manipulation of a topological space. The first step is to find a different way of defining the elements: "one way to improve the situation algebraically is to select an arbitrary point p in X and restrict our attention to paths which begin and end at p" (Crowell and Fox 1963, 13). Looking at loops (see Fig. 20) rather than simple paths provides a series of advantages:

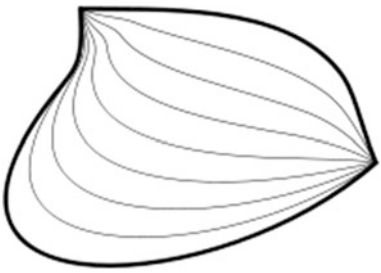


Fig. 19 A simple algebraic structure of a topological space

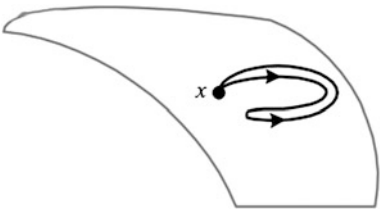


Fig. 20 A loop

1. if we can continuously deform a loop into another loop under homotopy, the we can consider them as a single element of our structure. Another loop that cannot be continuously deformed into the first is a different element. The bottom line: they are a means to tell apart specific region of the topological structure.
2. loops keep track of crucial properties of manifolds, such as holes in the surface.
3. It is easy to define a function of composition between loops. Since there will be a class of loops starting and ending at the same place, their composition (e.g. product of loops) is certainly defined at the base point.

Moreover the base-points act as the identity elements *i*—they are loops standing just at the base-point. Hence, the identity path *i* is a multiplicative identity. In this way we get a semi-group with identity on the topological space: the set of all *p*-based loops in *X*. This is better than the set of all paths of a topological space with respect to the product. But it can be improved. By introducing orientation for loops we can define the inverse of a loop and deepen the analogy with algebra: it is given by simply traversing it in the opposite direction. We are now in a position to build a group, as we have a *class of elements* (all the oriented paths starting and ending at a given base-point), a *product of paths* $(a*b)(t)$; the identity path *i*; the *inverse* of a path, denoted a^{-1}. The product is *associative*, that is $(ab)(c) = (a)(bc)$—but not *commutative*. This group for a topological space is called the *fundamental group* (see Fig. 21). The fundamental group is independent of the choice of a base-point: any loop through a point *p* is homotopic to a loop through any other point *q*.

This conceptualization, shaped by an analogy with algebra, gradually introduces information about topology that was not contained in it at the beginning of the process. It is the result of a new language and discourse about operations and elements of surfaces under homotopy. In this sense the fundamental group is a 'hybrid' (Grosholz 2007) and lays the foundation of algebraic topology: it is a basic tool for forming

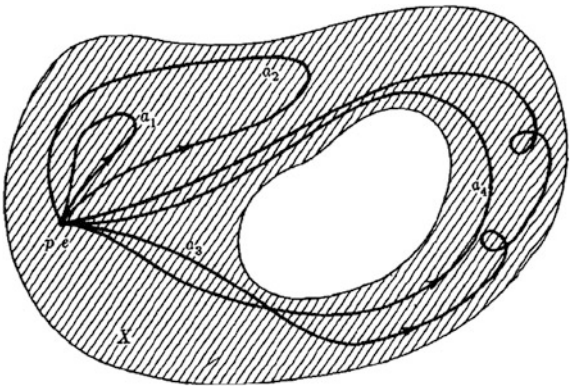


Fig. 21 The fundamental group

algebraic representations of topological spaces. Most often these algebraic representations are groups, but also structures like rings and modules can be associated to topological entities. The bottom line: algebraic topology «associates algebraic structures with purely topological, or geometrical, configurations. The two basic geometric entities of topology are topological spaces and continuous functions mapping one space into the other. To the spaces and continuous maps between them are made to correspond groups and group homomorphisms» (Crowell and Fox 1963, 13).

Now we can calculate the fundamental group π of several surfaces (sphere, anulus, torus, etc.) in order to tell them apart. Unfortunately, the fundamental group is not a complete invariant for 3-manifolds, and more generally «the algebra of topology is only a partial characterization of the topology» (Ibid.). More specifically, algebraic topology will produce one-way outcomes, which state that «if topological spaces X and Y are homeomorphic, then such and such algebraic conditions are satisfied. The converse proposition, however, will generally be false. Thus, if the algebraic conditions are not satisfied, we know that X and Y are topologically distinct. If, on the other hand, they are fulfilled, we usually can conclude nothing. The bridge from topology to algebra is almost always a one-way road» (Crowell and Fox 1963, 13). But even with this, we can do a lot.

5.1 Fruitfulness of Fundamental Group

The construction of an algebraic image of topological objects allows us to use algebraic tools and results as a means for investigating topological objects. Even if the problems that we can pose and solve, and the proprieties you can investigate, will depend on the adopted representation, we can generate new knowledge. For instance, the fundamental group allows us to solve problems and raise new problems.

On one hand, the Poincaré conjecture is a stock example of new problem that it generates. Its formulation is based on the notion of fundamental group: "consider a compact 3-dimensional manifold V without boundary. Is it possible that the fundamental group of V could be trivial, even though V is not homeomorphic to the 3-dimensional sphere?" (Poincaré 1906). This problem was solved by Perelman (2002, 2003a, b).

On the other side, the fundamental group can be used to solve problems, for instance just in knot theory. Wirtinger (1905) showed that trefoil is really knotted by proving that the fundamental group of the trefoil is the symmetric group on three elements. Moreover, Wirtinger extended his method so to construct the fundamental group of an arbitrary link. This presentation of the fundamental group is now known as the Wirtinger presentation. In addition, Dehn (1910) developed an algorithm for constructing the fundamental group of the complement of a link. He showed that a knot is nontrivial when its fundamental group is non-abelian, and that a trefoil knot and its mirror image are topologically distinct.

6 Remarks on Ways of Ampliating Knowledge

These examples reveal a number of crucial features about the rational and inferential ways whereby knowledge is ampliated. In particular, first I will examine the role played by representation, in particular the issues of the sensitivity to representation and the one of the interaction between representations of the same object (and their possible convergence, see also Morrison 2000 on this point). I will then examine the manipulation of mathematical objects, the role of ampliative reasoning, the historical nature of ampliation of knowledge, the role of theorem-proving and, in the end, the nature of mathematics objects.

First, both knot theory and topology exemplify the crucial role played by representation in the ampliation of knowledge. Here new knowledge and formal results come out from new representations that require specific manipulations of mathematical objects, in the sense of adapting or changing them to suit a given purpose or advantage. These manipulations end up with a conceptualization that produces a new language and discourses about a problem and its entities. The building of a new representation is a step-by-step process, which highlights certain features of the object and deliberately neglects other ones. This construction, when successful, creates an information surplus: introduces pieces of information that are not contained in the entities of a problem at the beginning of the process. Moreover, this construction is shaped mostly by ampliative inferences, as they are a means for progressively defining objects, relations, and the constraints[3] (or the conditions of solvability[4]) of a problem. For example the analogy with algebra is the key for

[3]See Nickles (1980a), and Nickles and Meheus (2009).

[4]See Cellucci (2013).

generating 'constraints' in the construction of the fundamental group. The generation of a new representation draws on existing knowledge, which is used in combination with ampliative inferences to suggest how to change or adapt a geometrical object to a specific purpose.

The role of representation in advancing knowledge raises the question of the sensitivity to representation. Since the construction of the representation highlights certain features of the objects and deliberately neglects other ones, the results derived from a new representation could be dependent on it. Let us go back to the simple use of the projection of a knot onto a plane: since several 3d features of knots are lost and other added (e.g. the overcrossings), we cannot be sure that results obtained for the 2d object will hold also for the 3d entity. We need the Reidemeister theorem to establish that a set of operations on 2d projections are valid also for the 3d knot. Moreover, if different representations reveal different properties of the object, then, in turn, we have another issue, namely the relation between these representations, in particular their convergence or not and, in the end, their possible unification. More specifically, we are interested in understanding if different properties revealed by different representations contribute together to the solution of a problem or not. Maybe they add pieces of the same puzzles, or maybe not. Moreover, even if they are adding pieces of the same puzzle, we do not know if they will be able to complete the picture—if their summing up will get a solution for the problem. They could never converge into a final solution to the problem, simply offering a "patchwork" (Cartwright 1999), a juxtaposition of characteristics of entities of the problem. Coloring, braids, numbering, graphs did not solve the problem of classification of knots, but offer an answer to portions of it, revealing partial class of equivalences.

In effect, the use of all these different approaches seems to tell a story of failure: the unsuccessful search for a solution to the problems of classification of knots and of manifolds. Each time, we construct and then pass from one approach to the other in search for a complete invariant that in the end we do not get. On the contrary, this seeming failure reveals many features of the ampliation of knowledge, displaying the several approaches employed to tackle a problem, how to switch from one to another and, when possible, to compose them. It would be much more difficult to reconstruct this process using a straight successful story of problem solving, where the right path to the solution is found from the very beginning. The absence of a solution, or at least the difficulties in finding it, force us to construct and put in use different tools, which to a large extent have to be made explicit. And so they can be more easily investigated. In the end, all this shows the multiplicative and manipulative nature of ampliation of knowledge. New representations, the multiplication of readings and viewpoints on the 'same' object, and ampliative inferences, are always in operation when we try to extend our knowledge.

Moreover the partial accounts to the problem of knots classification provided by the several approaches show an essential property of mathematical objects, as argued by Cellucci (2013). In essence these objects are inexhaustible, as new view-points can always be offered. I have shown how this happens via the construction of new representations. In this sense, an object is always partial and open

to new determinations. Or better, a mathematical object is simply a hypothesis put forward to solve a given problem and can always be conceptualized in new ways. Sometimes these ways converge and eventually merge into a deeper understanding of a problem and its entities, while sometimes they make sense only of certain features. In this sense also the distinction between a mathematical and a physical object almost vanishes. Knots are exemplary under this respect: tellingly, at the very beginning they were tackled with an experimental approach—an empirical compilation of tables of knots.

The examples in this paper show that the generation of new knowledge is not only a historical but also a holistic process.

Firs, it is historical in the sense that the new knowledge draws on the *corpus* of existing knowledge and since this varies trough time, the new knowledge that we can get will depend on it. Different configurations of this *corpus* will generate different pieces of new knowledge. So a problem that cannot be solved or posed a time t, could be so at $t + k$ when the *corpus* has changed. The production of new representations and viewpoints on problematic issues are based on the specific configuration of this *corpus*: the hypotheses candidate to solve certain problems and their order of introduction will change if it varies.

Second, it is holistic, both at interfield and intrafield level. As a part of our 'web' of knowledge expands, to use a long-standing metaphor, and new results emerge in a local area, they affect in principle any other area of our knowledge. Once a new tool or viewpoint is developed to solve a specific problem into a given field, it is open to the application to other problems in the field (intra-field) and also in another fields (inter-fields) via the construction of suitable representations. When this application succeeds, also the new domain will benefit from this application, by enlarging its corpus of knowledge. In turn, the knowledge produced in the new field can affect and expand the original source. For example, after the application of graphs to knot theory, we can go the other way by studying knotted graph, using knots to better understand and explore graphs (see e.g. Foisy 2002, 2003). Moreover, this holds also for different domains, for instance mathematics and physics.[5]

We have also seen that theorem-proving can play a heuristic role in our web of knowledge, exemplified by representation theorems like the Alexander's theorem. Here a new 'representation' establishes and formalizes a bridge between two mathematical entities and fields. In essence, theorem-proving creates new links in our web of knowledge by establishing a formal relation between two objects formerly treated as separated. Also re-proving theorem can be a heuristic move under this respect. Drawing the same consequence from different hypotheses reshape the web of knowledge, by connecting two previously unrelated areas and creating new links. Of course this is a different kind of 'new' knowledge.

More in general, the concept of novelty is positional, that is contextual and temporal: something is 'new' only with respect to a *specific body of knowledge at a*

[5]String theory is a stock example of the interaction between knot theory and physics.

given time, not by itself. So this body is the basis for the construction of any kind of new knowledge. This holds for virtually any ways of advancing knowledge, even the one based on the deliberate break of constraints (Nickles 1980a) or on the 'what if' approach (Nickles 2014). The first step of this construction, as we have seen, is the determination of similarities and dissimilarities between what we already know and the objects of our inquiry. In this process we employ an active manipulation of the target domain whereby new information is added to it and where we proceed step-by-step guided by analogies, and ampliative inferences in general.

This account of the construction of new knowledge faces two straight objections (e.g. Bunge 1981), namely:

1. analogies will collapse even if initially fertile, ending up from insight to nonsense;
2. rigorous interpretation and explanation are literal, not metaphorical.

According to this criticism, analogy plays only a logically and temporally limited role in the construction of knowledge. Juts like Wittgenstein's ladder, we have to throw it away after we have climbed up on it: "analogy can be fruitful in the preliminary exploration of new scientific territory, by suggesting that the new and unknown is, in some respects, like the old and known. […] whether or not the hypothesis succeeds, we shall have learned something, while nothing will have been learned if no hypothesis at all had been formulated" (Bunge 1981, 269). More specifically, if the analogy succeeds, "we shall know that A and B are indeed similar either substantially or formally. And if the analogy fails utterly, we shall realize that some radically new ideas are called for because B is in some respect radically different from A" (Ibid.). But even though analogy offers an epistemic gain in any case, it seems that it is what is radically new and *sui generis* that cannot be accounted for by means of analogies and ampliative inferences. The bottom line: analogy "on the one hand facilitates research into the unknown by encouraging us to tentatively extend our antecedent knowledge to a new field. On the other hand, if the world is variegated then analogy is bound to exhibit its limitation at some point, for what is radically new is precisely that which cannot be fully accounted for in familiar terms" (Ibid.). Analogy has to abandoned at a certain point, as it is just an approximation, but "unless the analogy is extremely specific or detailed, chances are that it will hold to a first approximation for, after all, our conceptual outfit is limited and no two concrete systems are dissimilar in every respect. The question is to decide what to stress at a given stage of research: whether resemblance or difference should be emphasized" (Ibid., pp. 268–9). Hence, analogy is simply an economical way of approaching new lands: "when faced with novelty we spontaneously start to oscillate between two poles: we try to find out what the new thing looks like and what it is unlike. Taking cognizance of resemblances and differences is the beginning of knowledge" (Ibid., 266).

But as we have seen in the case of knot theory and topology, analogy is not static transfer of knowledge, but a dynamic and active construction. More precisely it is not simply a transfer of a given and fixed knowledge from a source to a target, but it requires a dynamic process whereby the target has to be interpreted and

manipulated in order to make the transfer possible and obtain new knowledge. I have shown how during this process new information is gradually introduced in the target, displaying that 'new' is not an absolute concept—again, a piece of knowledge is 'new' always in comparison to something known and familiar. Any inquiry cannot even begin without something already known that we use as a basis for our investigation (searching for similarities and dissimilarities) and conceptualization, which can end up with a formal theory. In essence, other tools, theories and objects (the source in the analogical reasoning) suggest and shape the ways of manipulating the target under investigation. For instance, we have seen how graphs shape the way of manipulating and interpreting knots, i.e. by coloring the regions of the knot diagram alternately white and black, placing a vertex inside each white region, and then connecting vertices by edges going through the crossing points of the diagram.

More generally, this objection draws on a static conception of knowledge and the objects of scientific inquiry, whereby in principle objects can be determined in a complete fashion—they have a given, exhaustible set of properties that we can approach and ultimately grasp. On the contrary, in this paper I have shown the need for a dynamic and heuristic conception of knowledge and its objects in order to make sense in an appropriate and effective way of how we produce new knowledge. New representations offer a partial, open-to-new-determinations modeling of the objects of our inquiry and gradually allow us to construct and define goals, entities, operators, and the constraints in the problem-space of our inquiry. This construction is simply a human way of understanding and conceptualizing a phenomenon: it is not, and cannot be, literal (on this point see e.g. Lakoff and Johnson 1999), as it is always mediated by our biological and cognitive make-up. On second thought, it is even hard to see what a literal description is (see Turner 2005).

In the end, analogy and ampliative inferences are not simply the beginning of knowledge, but the only possible way of constructing knowledge and conceptualizing with it. Simply, we do not have other options. Novelty does not show up by chance, and the simple act of identifying a piece of knowledge as new is a step along a continual understanding that draws on them. The construction of knowledge relies on the search for similarities and dissimilarities (see Helman 1988): "analogies exist because of the way we categorize" (Turner 1988, 3), and they are "an indispensable, pervasive form of understanding" (Johnson 1988, 38). Even if 'new' is something whose characteristics are different from those of everything we already know, we must use what we already know to develop our understanding of it. Were this not the case, if something *totally* new came up (that is, something that cannot be reduced at least in part to what we know), no human way of understanding could make sense of it. It would be like facing the ocean on Solaris, an entity for which "the sum total of known facts was strictly negative" (Lem 1961, p. 23), and hence destined to remain unknown.

Acknowledgements I would like to thank Justin Roberts (Dept. Mathematics of University of California, San Diego) for his help with knot theory and topology.

References

Abbott, A.: Method of Discovery. W.W. Norton & Company Inc., New York (2004)

Adams, C.: The knot book. An Elementary Introduction to the Mathematical Theory of Knot. AMS, Williamstown (2004)

Alexander, J.W.: A lemma on systems of knotted curves. Proc. Nat. Acad. Sci. USA **9**, 93–95 (1923)

Alexander, J.W., Briggs, G.B.: On types of knotted curves. Ann. Math. Second Series, **28**(1/4), 562–586 (1926–27)

Artin, E.: Theory of braids. Ann. Math. 2nd Ser. **48**(1), 101–126 (1947)

Brunn, H.K. (1897). Über verknotete Kurven, Verh. Math. Kongr. Zürich, pp. 256–259

Bunge, M.: Analogy in quantum theory: from insight to nonsense. Br. J. Philos. Sci. **18**(4) (Feb., 1968), pp. 265–86 (1981)

Cartwritght, N.: The Dappled World. Cambridge University Press, Cambridge (1999)

Cellucci, C.: Rethinking Logic. Logic in Relation to Mathematics, Evolution, and Method. Springer, New York (2013)

Conway, J.H.: An enumeration of knots and links, and some of their algebraic properties. In: Leech, J. (ed.) Computation Problems in Abstract Algebra, pp. 329–358. Pergamon Press, Oxford (1967)

Crowell, R.H., Fox, R.: Introduction to Knot Theory. Springer, New York (1963)

Darden, L. (ed.): Reasoning in Biological Discoveries: Essays on Mechanisms, Inter-field Relations, and Anomaly Resolution. Cambridge University Press, New York (2006)

Dehn, M.: Über die Topologie des dreidimensionalen Raumes. Math. Ann. **69**, 137–168 (1910)

Foisy, J.: Intrinsically knotted graphs. J. Graph Theory **39**(3), 178–187 (2002)

Foisy, J.: A newly recognized intrinsically knotted graph. J. Graph Theory **43**(3), 199–209 (2003)

Gauss, F.C.: Disquisitiones Arithmeticae. Springer, Lipsia (1798)

Gillies, D.: Revolutions in Mathematics. Oxford University Press, Oxford (1995)

Grosholz, E.: Representation and Productive Ambiguity in Mathematics and the Sciences. Oxford University Press, Oxford (2007)

Grosholz, E., Breger, H. (eds.): The Growth of Mathematical Knowledge. Springer, Dordercht (2000)

Hanson, N.: Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science. Cambridge University Press, Cambridge (1958)

Helman, D.H. (ed.): Analogical Reasoning. Springer, New York (1988)

Ippoliti, E.: Between data and hypotheses. In: Cellucci, C., Grosholz, E., Ippoliti, E. (eds.) Logic and Knowledge. Cambridge Scholars Publishing, Newcastle Upon Tyne (2011)

Ippoliti, E.: Inferenze Ampliative. Lulu, Morrisville (2008)

Ippoliti, E.: Generation of hypotheses by ampliation of data. In: Magnani, L. (ed.) Model-Based Reasoning in Science and Technology, pp. 247–262. Springer, Berlin (2013)

Ippoliti, E. (ed.): Heuristic Reasoning. Springer, London (2014)

Johnson, M.: Some constraints on embodied analogical understanding. In: Helman, D.H. (ed.) Analogical Reasoning: Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy. Kluwer, Dordrecht (1988)

Kauffman, L.H.: Knots and Physics, Series on Knots and Everything—vol. 1. World Scientific, Teaneck (NJ) (1991)

Kauffman, L.H.: On knots. Ann. Math. Stud. vol. 115. Princeton University Press, Princeton (1987)

Kinoshita, S., Terasaka, H.: On Unions of Knots. Osaka Math. J. **9**, 131–153 (1957)

Lakatos, I.: Proofs and Refutations: The Logic of Mathematical Discovery. Cambridge University Press, Cambridge (1976)

Lakoff, G., Johnson, M.: Philosophy in the flesh. Basic Books, New York (1999)

Laudan, L.: Progress and its Problems. University of California Press, Berkeleyand LA (1977)

Lem, S.: Solaris. Faber & Faber, London (1961)

Listing, J.B.: Vorstudien zur Topologie, Gottinger Studien (Abtheilung 1) 1, 811–875 (1847)
Magnani, L.: Abduction, Reason, and Science. Processes of Discovery and Explanation. Kluwer Academic, New York (2001)
Magnani, L., Carnielli, W., Pizzi, C. (eds.): Model-Based Reasoning in Science and Technology: Abduction, Logic, and Computational Discovery. Springer, Heidelberg (2010)
Mazur, B.: Notes on ´etale cohomology of number fields. Ann. Sci. Ecole Norm. Sup. **6**(4), 521–552 (1973)
Montesinos, J.M.: Lectures on 3-fold simple coverings and 3-manifolds. Contemp. Math. **44** (Combinatorial methods in topology and algebraic geometry), 157–177 (1985)
Morrison, M.: Unifying Scientific Theories. Cambridge University Press, New York (2000)
Nickles, T. (ed.): Scientific Discovery: Logic and Rationality. Springer, Boston (1980)
Nickles, T., Meheus, J. (eds.): Methods of Discovery and Creativity. Springer, New York (2009)
Nickles, T.: Heuristic appraisal at the frontier of research. In: Ippoliti, E., (ed.) Heuristic Reasoning, pp. 57–88. Springer, Milan (2014)
Perelman, G.: Ricci flow with surgery on three-manifolds. arXiv:math.DG/0303109 (2003a)
Perelman, G.: Finite extinction time for the solutions to the Ricci flow on certain three-manifolds. arXiv:math.DG/0307245 (2003b)
Perelman, G.: The entropy formula for the Ricci flow and its geometric applications. arXiv:math.DG/0211159 (2002)
Poincaré, H.: Analysis situs. J. de l'École Polytechnique. **2**(1), 1–123 (1895)
Poincaré, H.: The present and the future of mathematical physics. Bull. Amer. Math. Soc. **12** (5), 240–260 (1906)
Polya, G.: Mathematics and Plausible Reasoning. Princeton University Press, Princeton (1954)
Reidemeister, K.: Knotten und Gruppen. Abh. Math. Sem. Univ. Hamburg **5**, 7–23 (1927)
Reidemeister, K.: Knotentheorie. Julius Springer, Berlin (1932)
Schubert, H.: Knoten mit zwei Brucken. Math. Z. **65**, 133–170 (1956)
Schubert, H.: Die eindeutige Zerlegbarkeit eines Knotens in Primknoten. S.-B Heidelberger Akad. Wiss. Math. Nat. Kl. **3**, 57–104 (1949)
Simon, H.: Models of Discovery. Reidel, Dordrecht (1977)
Simon, H., Langley, P., Bradshaw, G., Zytkow, J. (eds.): Scientific Discovery: Computational Explorations of the Creative Processes. IT Press, Boston (1987)
Tait, G.: On knots I, II, III. Scientific Papers, vol. 1, pp. 273–347. Cambridge University Press, London (1900)
Tait, P.G.: Some elementary properties of closed plane curves. Messenger of Mathematics, New Series, n. 69. In: Tait. G. (ed.) (1898) *Scientific Papers*. vol. I. Cambridge University Press, Cambridge (1887)
Thomson, W.H.: On vortex motion. Trans. Roy. Soc. Edinburgh **25**, 217–260 (1869)
Turner, M.: Categories and analogies. In: Helman, D.H. (eds.) Analogical reasoning: perspectives of artificial intelligence, cognitive science, and philosophy. Kluwer, Dordrecht (1988)
Turner, M.: The literal versus figurative dichotomy. In: Coulson, S., Lewandowska-Tomaszczyk, B. (eds.) The Literal and Nonliteral in Language and Thought, pp. 25–52. Peter Lang, Frankfurt (2005)
Weisberg, R.: Creativity: Creativity: Understanding Innovation in Problem Solving, Science, Invention, and the Arts. Wiley, Hoboken (NJ) (2006)
Wirtinger, W.: Uber die Verzweigungen bei Funktionen von zwei Veriinderlichen. Jahresbericht d. Deutschen Mathematiker Vereinigung **14**, 517 (1905)
Nickles, T. (ed.): Scientific Discovery: Case Studies. Springer, Boston
Yajima, T., Kinoshita, S.: On the graphs of knots. Osaka Math. J. **9**(2), 155–163 (1957)

Models, Idealisations, and Realism

Juha Saatsi

Abstract I explore a challenge that idealisations pose to scientific realism and argue that the realist can best accommodate idealisations by capitalising on certain modal features of idealised models that are underwritten by laws of nature.

1 Introduction

This paper explores a challenge that idealisations pose to scientific realism. I will review the challenge before briefly assessing some recent analyses of idealised models that function as a foil and motivation for my response to the challenge. I will argue that the realist can best accommodate idealisations by capitalising on certain modal features of idealised models that are underwritten by laws of nature.

The idea that idealisations in some sense represent *possibilia* is common place. Typical idealisations—such as frictionless planes, point masses, isolated systems, and omniscient agents—are naturally thought of in terms of possible systems that are suitably related to some actual systems of interest. David Lewis, for example, thought that we can best make sense of the pervasive utility of idealisations in science in terms of possible worlds that are more and less similar to the actual world.

> [We find it much easier to tell the truth if we sometimes drag in the truthlike fiction, and when] we do, we traffic in possible worlds. Idealisations are unactualised things to which it is useful to compare actual things. An idealised theory is a theory known to be false at our world, but true at worlds thought to be close to ours. The frictionless planes, the ideal gases, the ideally rational belief systems—one and all, these are things that exist as parts of other worlds than our own. The scientific utility of talking of idealisations is among the theoretical benefits to be found in the paradise of possibilia. (1986, pp. 26–27)

J. Saatsi (✉)
School of Philosophy, Religion, and History of Science, University of Leeds, Leeds, UK
e-mail: J.T.Saatsi@Leeds.ac.uk

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_10

Recognising that idealisations are naturally thought of in terms of possibilia is but a start, however. The spadework lies in properly accounting for the utility of such modal constructs.[1] What, then, is required to account for the utility of idealisations in science? Various questions present themselves here. I will focus on the following challenge to scientific realism, in particular: given that idealisations incorporate false assumptions about the way the world is, why are idealisations so important for making successful predictions and for coming up with powerful explanations?

I will examine this challenge in relation to idealised *models* in particular. (Often theoretical *laws* are also characterised as idealised. My focus is on idealised modelling assumptions other than laws.) This challenge differs from the standard arguments against scientific realism, deriving from a 'pessimistic induction' over past false theories, or the idea that theories can be underdetermined by evidence. These stock anti-realist arguments are typically framed in terms of scientific *theories*. It is interesting that the realism debate has been largely framed in terms of theories, even though in contemporary philosophy of science much of the focus has shifted from theories to models as the most fitting 'unit' of philosophical analysis. In as far as realism is primarily motivated by the impressive empirical success of science (culminating in novel predictions), it is typically models that provide or facilitate such success. Furthermore, according to a popular 'modelling view' of science, theories in a sense are nothing but families of models (unified by laws).

When we shift the focus from theories to models, anti-realists can find further ammunition from various kinds of inconsistencies that modelling practices exhibit. Often different models of one and the same phenomenon are mutually inconsistent with one another. Some models are even internally inconsistent. And many models incorporate assumptions that are at odds—sometimes radically so—with modellers' background beliefs. Such inconsistencies can be used to challenge the realist in as far as they indicate that various kinds of falsehoods are playing a bigger role in the production of the empirical successes that realists are inclined to think. If falsehoods can play a significant role in bringing about empirical successes, perhaps the role played by (approximate) truths is less significant than realists would have it? Perhaps the joint contribution to empirical success from idealising falsehoods, and whatever degree of (approximate) truth there is to a model otherwise, can be so entangled that we can make no sense of the realist credo that a model's empirical success is *due to* its 'latching onto reality'?

What follows is concerned with this kind of challenge, arising out of the indispensability of idealisations for modelling. I will start by fleshing out the challenge (§2), before briefly reviewing some philosophical analyses of idealisations (§3), paving the way for my own response to the challenge (§4).

[1] A lot has been written about modal aspects of idealisations. I will not attempt to relate my point of view here to the broader context of the Poznań school and the verisimilitude literature, for example. See Niiniluoto (2007) for a review.

2 Realism and Idealisations: A Challenge

Let us first try to get a good handle on the 'idealisation-challenge' previewed above. How exactly do idealised models challenge a realist attitude to science? Sorensen (2012) crisply (if somewhat provocatively) explains:

> Scientists wittingly employ false assumptions to explain and predict. Falsification is counter-productive in the pursuit of truth. So scientific realism appears to imply that idealisation would be worse than ineffective.
> The instrumentalist says the scientist merely aims at the prediction and control of phenomena. […] Given that scientists are indifferent to the truth and often believe idealisations will promote prediction and control, the instrumentalist predicts that the scientists will idealise.
> Consequently, idealisation looks like a crucial experiment for philosophy of science. […] Since scientists idealise, the instrumentalism prevails. (p. 30)

In other words, if realism is committed to the notion that science aims at truth, while anti-realists regard lesser aims of empirical and instrumental adequacy to be enough, then idealisations seem to speak against realism.

One might worry that this challenge to realism quickly evaporates in the light of obvious realist responses applicable to many (or perhaps even most) idealisations.[2] Consider various 'Galilean' idealisations, for example, that McMullin (1985) views as providing an argument *for* scientific realism, not against it. Take an idealised model of a gravitational pendulum, for instance. It incorporates various simplifying assumptions, such as the complete absence of air resistance, friction, and so on. But it does so in a way that readily suggests ways in which the model can be *de-idealised*, for example by simply adding further terms to the model's force function. McMullin rightly points out that a realist reading of idealised models best predicts and explains models' capacity to be thus de-idealised; therefore such idealisations arguably support (a suitably qualified form of) realism about these kinds of idealised models.[3]

In general, a realist perspective on scientific modelling clearly has the wherewithal to account for the way in which various kinds of simplifications are

[2]To be clear, Sorensen himself notes that idealisations only 'appear' to challenge scientific realism, and he does not endorse the instrumentalist conclusion in the offing. I will review Sorensen's reasoning in §3.

[3]See McMullin (1985):

> If the original model merely 'saved the appearances' without in any way approximating to the structure of the object whose behavior is under scrutiny, there would be no reason why this reversal of a simplifying assumption, motivated by the belief that the object does possess something like the structure attributed to it, would work as it does. Taking the model seriously as an approximately true account is what leads us to expect the correction to produce a verifiable prediction. The fact that formal idealisation rather consistently does work in this way is a strong argument for a moderate version of scientific realism. (p. 262)

pragmatically indispensable in the scientific study of systems of otherwise unmanageable complexity. Realist reading of theories suggests different ways of brushing aside complications that, according to our theory, make next to no contribution to the end result. After all, science is obviously *not only* in 'the pursuit of truth', even according to the realist; it is also in the pursuit of achieving actual results, mathematical tractability, predictions, effective control and manipulability, and so on. The different aspects of Galilean idealisations are ways in which the realist can anticipate deliberate 'falsifications' (typically simplifications) to contribute to the latter pursuits.[4]

This realist response does not answer the challenge completely, however, since some idealisations do not fit the Galilean mould. Philosophers of science have identified other, more radical idealisations in science, and the question remains whether some of these non-Galilean idealisations rather support instrumentalism about certain kinds of models. What should a realist say about 'uncontrollable' idealisations where no de-idealisation is in the offing? (see e.g. Batterman 2005) What about idealisations involved in the so-called minimal models, such as the Ising-model? How should the realist accommodate these kinds of idealisations that seem altogether indispensable, going beyond the kind of broadly pragmatic convenience associated with Galilean idealisations? How can the realist account for the indispensable utility of such falsifications in modelling? This is one challenge that remains for the realist.

Moreover, even with respect to Galilean idealisations, there is further work to be done in clarifying the letter of the realist response. For example, is there a conceptual framework within which the utility of different types of idealisations can be accounted for in unified terms? Intuitively speaking, the realist response to the challenge from idealised models is to say that there is a sense in which an idealised model 'latches onto' reality in a way that is responsible for the model's empirical success. One challenge is to articulate this notion of 'latching onto reality' so as to capture the relevant features of models in a way that meshes with the realist intuitions. Call this the articulation-of-realism challenge. What does an idealised model 'get right' about its target system, such that it is empirically successful *by virtue of* getting those things right (and despite getting some other things wrong)? This challenge of articulating how idealised models latch onto reality has been recognised in the vast literature on idealisations, and realists typically maintain that there is some principled sense in which predictive (as well as explanatory) success is due to models latching onto reality. This then underwrites the realist's epistemic commitment for regarding predictive success as a (fallible) indicator of models latching onto reality in this sense.

Philosophers have appealed to different conceptual and formal resources in spelling out this idea, ranging from accounts of verisimilitude, to partial structures and quasi-truth, to philosophy of language/logic, to philosophy of fiction.[5] I cannot

[4]McMullin (1985) distinguishes three different types of Galilean idealisations.

[5]See e.g. Niiniluoto (2007), Da Costa and French (2003), Sorensen (2012), Toon (2012).

do full justice to this rich literature here, but I will next briefly review a couple of recent analyses of idealisation as a foil for my own perspective. To prefigure: in my view these analyses fall short of properly accounting for the empirical success of idealised models. Providing a sense in which a model can get things right, while also getting things wrong, does not in and of itself account for how the falsehoods are *immaterial* for the empirical successes at stake, and how the empirical successes are *due to* 'getting things right'. After discussing these analyses of idealisations I will take steps towards a different (possibly complementary) account of idealisations that better serves the realist's need to explain the empirical success of idealised models. This requires reflecting more closely on what it takes to account for predictive success of a model that incorporates false assumptions. I will argue that such an account can turn on showing how a model's predictive success is *robust* with respect to variation in the false assumptions involved in idealisations, in the sense that these assumptions could have been different without undoing the predictive success. I will argue that it is this modal character of idealisations that can account for their utility from a realist perspective.

3 Some Analyses of Idealised Models

Idealisations as suppositional. Recall Sorensen's presentation of the idealisation-challenge above. His own response to it is iconoclastic. Typically philosophers characterise idealisation as being essentially a matter of some sort of intentional introduction of distortion into a scientific model or theory, with different philosophers holding different views regarding the nature of such 'intentional distortions'. For example, such intentional distortion has been taken to be a matter of *indirect* assertion of something true (Strevens); *relativized* assertion of something true (Giere); *temporary* assertion of falsehood (McMullin); or assertion in the mood of *pretence* (Toon, Frigg, and various others). In contrast to these different ways of regarding idealisation as some sort of attenuated assertion, Sorensen views idealisations as *suppositional*, in analogy to suppositional premises in a conditional proof.[6] That is, Sorensen's perspective on idealisation—drawing on philosophy of language and logic—regards it as a matter of 'simplifying supposition,' naturally free of any realist commitment. (Compare: a mathematician's supposition that 'there is a largest prime' for the purpose of *reductio ad absurdum* entails no commitment to finitude of primes.) In sum:

[6]Schematically: [(P1)] Suppose P.
[(P2)] From P derive Q.
[————————]
[(C)] Conclude that if P then Q.

> Idealisation is constituted by supposition. Only simplified suppositions count as idealisations. The filters are psychological and methodological. Idealisers seek tractability, memorability, and transmissibility (rather like myth makers). (Sorensen 2012, p. 37)

Sorensen contends that we can thus assimilate idealisations with something that is already well understood by logicians—a supposition that initiates a conditional proof or reductio ad absurdum. Allegedly we thus have an 'off-the-shelf' model for analysing idealisations as a matter of propositions that are governed by well-understood rules of rational use; not ontologically committing, for well-understood reasons; and not in need of elimination. And all this arguably explains, at least in part, why scientists are so happy to idealise, and are not overly preoccupied with de-idealisation or the 'distance' of verisimilitude between idealisation and the exact truth.

While Sorensen's perspective may throw light on some idealisations in science, for various reasons I do not find the analogy convincing or illuminating in general. Even the simplest paradigmatic exemplars of Galilean idealisation, such as the ideal pendulum or a frictionless plane, seem to be fit-for-purpose for obvious reasons that have little to do with *reductio ad absurdum*, or purely conditional arguments. Analysing idealisations in terms of the status of propositions involved—suppositional vs. assertoric—also seems much too dichotomous and coarse-grained to capture relevant differences in the various kinds of idealisations and how they contribute to predictive success (see e.g. McMullin (1985) for useful distinctions amongst different flavours of 'Galilean' idealisations, and Batterman (2005) for the distinction between these and 'non-Galilean' idealisations). Furthermore, with respect to the articulation-of-realism challenge most importantly, it is wholly unclear why a realist account of a model's predictive success should in any way depend on whether or not the 'falsehoods' involved are intentional, as in the case of idealisations, or simply mistaken assertions about the target. In both cases we can consider the relationship between the target-as-represented-by-the-model, and the target-as-it-actually-is, in trying to account for the model's empirical success in terms of how it latches onto reality.[7]

Idealisations in the semantic view. According to the semantic view of theories idealisations are more of a piece with other approximations. The semantic view is touted as providing a unified account of science where models occupy a centre stage. Consider da Costa and French (2003), for example, who offer models in the sense of (quasi-formal) model-theory as an appropriate backbone to a 'unitary approach to models and scientific reasoning.' In particular, their model-theoretic meta-scientific framework is motivated as offering the wherewithal to capture idealisations and approximations in science by providing 'a more sophisticated concept of 'model' […] which accommodates the essential incompleteness and partial nature of scientific representations' (p. 5). In their 'partial structures' formalisation of the semantic view, idealisations (as well as other approximations) can be 'accommodated through the introduction of 'partial isomorphism' as the fundamental relationship—horizontally and vertically—between theoretical and data

[7]The realist faces the epistemic challenge of justifying her knowledge of the target-as-it-actually-is, of course, but this issue has nothing to do with idealisation per se.

models' (p. 102). Furthermore, the model-theoretic framework furnishes a notion of 'quasi-truth' that 'can be used to formally underpin the claim that idealisations are regarded *as if* they were true' (p. 163). (See da Costa and French (2003) for details.) Moreover, arguably the considerable flexibility of the partial structures framework allows it to also accommodate non-Galilean infinite idealisations (Bueno and French 2012). It is thus offered as a truly unitary approach to understanding the role and workings of idealisations—both Galilean and non-Galilean alike.

Is it enough for a realist to point to this meta-scientific framework as providing a satisfactory response to the challenges that idealisations pose to her? I do not think so. The framework of partial structures, partial homo-/isomorphisms, and quasi-truth allows us to identify a formal correspondence between an idealised model and its target, which in turn allows us to formally (re)present the idea that the model is in a sense 'latching onto' the target. Since idealised models can latch onto their targets in this sense, the framework thus 'accommodates' idealisations. But we should try to go beyond this by *accounting for* an idealised model's empirical success by showing how a model's 'latching onto' unobservable reality can be considered to be *responsible* for the model's predictive success. It is not clear how the existence of partial homo-/isomorphisms between (a formal representation of) a model and its target, or the model's quasi-truth for that matter, provides understanding of why the model is empirically successful *by virtue of* latching onto reality thus-and-so, and *regardless of* incorporating such-and-such aspects of misrepresentation. We should want a clearer sense of the role played by the idealising 'falsehoods' in an idealised model, and a clearer sense of how the realist can bracket those aspects of the model as falling outside her realist commitments, despite them being useful, or even indispensable for making the predictions. The existence of partial homo-/isomorphism between an idealised model and a data model, for instance, says nothing about this in and of itself, and little has been said by way of analysing the explanatory credentials of such structural relations (vis-à-vis the idealisation challenge) in the context of the semantic view.[8]

Many have taken to heart the notion, well expressed by Giere, that when it comes to science 'idealisation and approximation are the essence, [so] an adequate theory of science must reflect this fact in its most basic concepts' (Giere 1988, p. 78). But in the face of the idealisation challenge 'reflecting' is not enough. An adequate (realist) theory of science should also *account* for the empirical success of idealised models, and the above accounts of idealisations fall short of throwing sufficient light on the roles played by idealisations in the production of predictive success. In particular, we should demand a clearer sense of how the realist can consider idealisations not to be the driving force behind models' predictive success, and how the realist can rather consider the models' latching onto reality to be responsible for it.

[8]It is possible that more can be said on behalf of the structuralist analysis of idealisation, and the partial structures analysis of idealisations can well be a useful part of a bigger picture, of course.

4 Towards a Realist Analysis of Idealisations

Scientific models and their inexact representational fit to the world raise various questions, many of which specifically concern idealisations. But it is important to realise that the articulation-of-realism challenge, in particular, is actually *not specifically* about idealisations. Rather, it is an instance of a much broader challenge to realism. The general form of the question at stake is: how can a model that is false in *this* way be empirically successful in *that* way? This question arises in connection with any empirically successful model that incorporates falsehoods, regardless of the reason behind those falsehoods. A model can incorporate falsehoods due to being idealised, but also for other reasons. In particular, the same question arises even if scientists are simply mistaken or misguided about their target of theorising.

Recognising the general nature of the question at stake, it is immediately unclear why the realist response to it should vary depending on the reason behind the representational inaccuracy in play. Why would it matter for the realist response whether the reason behind a representational inaccuracy is an intentional simplification (as in the case of idealisation), or an unintentional, erroneous assumption?[9] After all, in both cases the realist hopes to be able to answer this question in terms of how the model relates to its target, in such a way that we can regard the sense in which the model latches onto its the target as being responsible for the model's empirical success. Furthermore, if we have a fruitful conceptual framework for offering a realist response in connection with unintentional misrepresentations in science, it is reasonable to try to apply that framework also to idealisations (qua intentional misrepresentations).

My analysis of idealisations from a realist perspective is guided by this line of thought. That is, I adopt a conceptual framework that I have found fruitful and apposite in connection with some models that incorporate fundamentally misguided assumptions.[10] I claim that within this framework we can in quite general terms naturally account for idealisations' utility in modal terms, going beyond merely noting that idealisations traffic in non-actual possibilia to which actual systems can be usefully compared, or that there are different (quasi-formalisable) senses in which idealised models can be 'partially true' despite the 'falsehoods' they incorporate.

Here is an outline of the conceptual framework. We shall focus on predictive success of models, ignoring their explanatory success for now. (I will comment on the explanatory utility of idealisations later.) Given a particular model, I am

[9]There are questions about the modelling practice that specifically involve idealisations: for example, is the endemic and carefree employment of idealisations in tension with realism? My way of framing the idealisation-challenge focuses on models themselves, not the modelling practice.

[10]Some of these models have animated much discussion in the realism debate, such as Fresnel's elastic ether model of the partial refraction and reflection of light, used to derive the so-called Fresnel's equations. See Saatsi (2005).

interested in question Q: how is the model predictively successful—viz. empirically adequate in the relevant ways—despite misrepresenting its target in certain respects. To answer this question we can consider a range of models that vary in those respects, corresponding to a range of possible systems they can be taken to represent. The aim is to show how the required degree of empirical adequacy is independent of the particular false assumptions incorporated in the model. Independence is a matter of robustness of predictive success under variation in possible modelling assumptions that together with fixed background assumptions (including the relevant laws) yield the predictive success at stake.

The thought is that modal information of this sort can furnish a realist response to Q to the following extent: it shows how modelling assumptions can be false but nevertheless 'contain' veridical assumptions about the target that are responsible for the predictive success in the sense that variation in the specific false assumptions, without variation in the 'contained' veridical assumptions, would not undo the predictive success. It is via these 'contained' veridical assumptions that the model can be viewed as latching onto reality so as to ensure the model's predictive success. The specific false assumptions involved, regardless of their indispensability or otherwise for presenting and working with the model, are not doing any of the heavy lifting in producing the predictive successes at stake.

The tricky business lies in spelling out the sense in which a set of modelling assumptions can 'contain' veridical, success-fuelling assumptions. (The complex literature on verisimilitude and approximate/partial truth demonstrates how difficult these issues are.) Here I take 'containing' to be a matter of the specific modelling assumptions together with the relevant background assumptions entailing some further, less specific features of the target system, such that getting these further features right (in conjunction with the relevant background assumptions, including laws) would suffice for a model to exhibit the predictive success at stake.[11] All this is perhaps best elaborated by illustrating it via a simple toy example. Before we get to this, I note again that nothing in the abstract outline above directly corresponds to the notion of idealisation. This is as it should be, for the reasons given at the start of this section.

As for a toy example, consider a model system with a graph-like structure. The model represents its target system as having four nodes, connected by some dyadic relations as in Fig. 1.

That is, the model represents a target of four vertices, connected with one another in this 3-regular way. (A graph structure is 3-regular if each of the nodes is connected to three other nodes.) Assume that the relevant background assumptions, including the relevant laws, allow one to make a successful prediction about the system's behaviour under some circumstances (e.g. in the chemistry of carbon molecules.)

[11]By 'entailing' I mean not only logical entailment, but also metaphysical entailment, such as the relationship between determinate and corresponding determinable properties. If facts about such relationships can be packed into the background assumptions we can ensure logical entailment, of course.

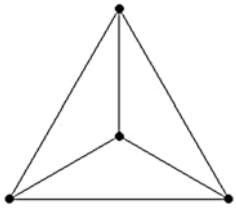


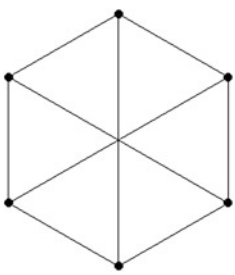
Fig. 1 3-regular model with four vertices

Assume further (for the sake of the argument) that the phenomenon in question, given the laws, is only exhibited by systems that have less than eight nodes, and that for such systems the phenomenon only depends on 3-regularity. That is, we are assuming that the relevant laws are such that even if the system were to have six nodes, say, and a 3-regular structure, it would display the behavior predicted.

Given these assumptions, representing correctly the number of nodes is clearly not relevant for the predictive success of our model. If as a matter of fact the target is 3-regular and has six nodes (as in Fig. 2), then the model misrepresents the target regarding the number of nodes and relations between them, but it still 'gets right' the fact that each of the nodes is connected to three other nodes, i.e. the structure is 3-regular.

I take it that there is an intuitively clear sense in which our model gets the relevant feature of the target right: it latches onto reality by correctly representing the target's 3-regularity. This is the critical, less specific feature of the system that the model 'contains'. It is less specific than the modelling assumptions that specify which node is connected to which. (Note that 3-regularity need not be part of the stated modelling assumptions, and need only be 'contained' in these assumptions in the sense of being entailed by them.) And it can be this sense of 'containing' of the veridical assumption about the target—this sense of 'latching onto' the target—that explains the model's empirical adequacy vis-à-vis the phenomenon in question. The model's predictive success is explained in a way that renders wholly immaterial the misrepresentations the model incorporates with respect to the number of nodes, and which node is connected to which. In the setting of this toy-example, grasping this sense 'latching onto' the target adequately answers the challenge at stake.

Fig. 2 The actual 3-regular target with six vertices



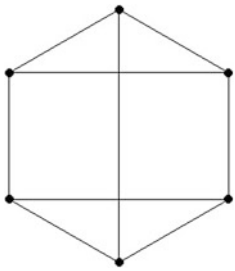


Fig. 3 The other possible 3-connected model with six vertices

For the very same reason a model that represents the target as a different 3-regular graph of six nodes (as in Fig. 3) would be equally empirically successful. It also 'gets right' the fact that the target is 3-regular, and that it has less than eight nodes. As it happens, there are only three 3-regular graphs of less than eight nodes. A given target system can only instantiate one of these specific structures, but given the laws (we have assumed), the relevant features that our model needs to latch onto are less specific than that. The critical, less-specific features of the target are: the target has 4 or 6 nodes; the target is 3-regular. These less-specific modelling assumptions are realized in three different, more specific ways. Any model is going to incorporate one or another of the specific realisers, but all that really matters is that a model incorporates *one or another* of these features, i.e. that it incorporates the less specific feature. Some models can furthermore count as being idealised by virtue of incorporating such specific realiser that sufficiently simplifies the model in its presentation and operation. (In some context a 3-regular graph of mere 4 nodes could be an idealisation of a larger 3-regular graph, for example.)

This is merely a simple toy example, of course, but it serves to bring out the key features of an interesting conceptual framework. In particular, it shows how accounting for a model's predictive success can turn on grasping the robustness of predictive success under variation in the specific modelling assumptions that all 'contain' a critical veridical assumption.[12] The sense in which a model (in relation to the relevant background assumptions) can thus latch onto reality is conceptually quite straightforward, and not in my view well captured by the existing (quasi-) formal frameworks for 'partial truth', approximate truth, or verisimilitude. What matters is the grasping of what is common to different possible systems, such that the common feature is all that matters, since variation in other features is immaterial: any model that features some 'realiser' of the common feature would count as predictively successful. A derivation of the prediction further requires the right laws of nature as background assumptions, grounding the relevance of these less-specific features.

[12]This has connotations of robustness analysis of idealised models (see e.g. Odenbaugh 2011). Exploring the connections to the literature on robustness analysis requires further work. (Thanks to Arnon Levy for flagging this question for me).

The sense in which modelling assumptions can 'contain' a veridical, success-fuelling assumption need not be captured by a notion of partial truth applied to propositions that can be used to specify the model. Consider, for example, the model:

{Alice knows Bob. Bob knows Erik and Fred. Erik knows David and Charlie. David knows Fred and Alice. Fred knows Charlie. Charlie knows Alice.}

This model can latch onto the target represented by

{Alice knows Bob. Bob knows Erik and Charlie. Erik knows David and Fred. David knows Alice and Fred. Fred knows Charlie. Charlie knows Alice.}

The two systems represented by these two sets of propositions exhibit the two alternative 3-regular structures with six nodes. Neither set of propositions explicitly says anything about the shared 3-regularity, however, and the underlying similarity is not explicitly represented by the propositions, nor revealed by looking at the (partial) truth or otherwise of the (sets of) propositions involved in presenting the two systems. Since the pertinent similarity between the model and the target need not be part of the explicit representational content of the model—the model need not represent the target *as* 3-regular—I call the model *inferentially* veridical (as opposed to representationally veridical). The idea is that from the model we can infer, with the help of the relevant background assumptions, the critical veridical assumptions.[13]

One may worry that this kind of 'inferential veridicality' is too thin to support a realist account of empirical success. One may worry, for example, how the less specific feature 'having 4 or 6 nodes'—a disjunctive property—can be attributed to the target. Or one may worry about the sense in which a model 'containing' a veridical assumption of this kind can account for the model's empirical adequacy in a realist spirit. I think the right response to such worries is to note that it is the appeal to laws of nature in deriving predictions from a model that underwrites the significance of the less-specific properties, regardless of whether or not they have disjunctive realisations. So, given these laws, from a scientific point of view such a property can be a genuine, bona fide feature of the world on which our theorising can latch, despite its disjunctive (or unspecific, or vague) character. One way to put this is to say that with the less-specific, veridical assumptions we are latching onto an important modal truth: had the target had only 4 (as opposed to 6) nodes, all with 3 connections, the same result would have ensued given the relevant laws of nature.

One may push the same worry in more general terms, in relation to my characterisation of how the veridical assumptions are 'contained' in the model. I said above that 'containing' is a matter of the specific modelling assumptions together with the relevant background assumptions *entailing* some further, less specific features of the target system. The worry here is that this idea that the model is thus

[13]The realist can then claim that derivations of successful predictions involve such inferences, and thus involve the veridical assumptions. Cf. Saatsi (2005) for related discussion in connection with Fresnel's model of light.

latching onto some less specific, more abstract worldly features seems to face a 'disjunction problem': since any modelling assumption p always entails $p \vee q$, any model is (allegedly) guaranteed to latch onto reality, as long as there is *some* true q such that it would work to produce the right prediction.[14] Does a model's inferential veridicality thereby become a trivial matter, deflating realism of any worthwhile commitment?

The answer is no. It is *not* the case that any model is guaranteed to latch onto reality just by virtue of being predictively successful, since a model latches onto reality partly by virtue of appealing to appropriate facts about laws of nature. For example, if one constructs an empirically adequate model M in classical Newtonian physics of a purely quantum phenomenon, the false modelling assumptions are not latching onto reality, since there is no possible classical model that provides a faithful, veridical representation of the target. It is not the case that some more complicated classical model faithfully represents the system and shares the critical, less specific properties with M such that any classical model that exhibits those properties would be equally empirically adequate as M. For the same reason a Ptolemaic model with epicycles does not latch onto its target (the solar system) despite its impressive empirical success.

Admittedly there is much more to be said regarding the kind of realism that can be served by the conceptual framework I am proposing here, and I hasten to add that it is not the case that realist intuitions and cause are saved *just* by showing predictively successful models being inferentially veridical. There can be interesting cases of local underdetermination where radically different modelling assumptions, in conjunction with the right laws, give rise to more or less the same predictions (see for example Saatsi and Vickers (2011) for one such case). In such cases the explanation of predictive success can have a strong anti-realist flavour. But in many cases the details of the derivation, and in particular the role played therein by the relevant less-specific features (with respect to which the model is inferentially veridical), can serve the realist cause by saving the 'no miracles' intuition. Or so I contend.

5 Beyond Toy Examples

I have proposed, largely in the abstract, a conceptual framework for accounting for the predictive success of idealised models in modal terms. One may wonder whether this conceptual framework can capture some real idealised models as well. I certainly think so! Consider a paradigmatic Galilean idealisation, such as an ideal pendulum as a model of my grand father's pendulum clock. The model's degree of empirical adequacy is naturally accounted for in terms of the model's inferential

[14]See Strevens's (2008) discussion of the disjunction problem in connection with his difference-making account of causal explanation that operates by abstraction.

veridicality, in conjunction with the appropriateness of the background laws (Newtonian mechanics + gravity). The model is inferentially veridical by virtue of entailing truths about less specific features of the target such that any model that realises those features in one way or another will attain at least that degree of empirical adequacy. The relevant less specific features concern a vague force function, vague specification of the pendulum's dimensions, etc. The ideal pendulum model represents a particularly simple specific realisation of these less specific (vague) features, and its empirical adequacy is easily accountable—regardless of its misrepresentation in these respects—by noting the robustness of its predictive success under variation in the particular false specification of the critical less specific features, the specification that constitute the idealisation.

Various other Galilean idealisations similarly lend themselves to analysis in these modal terms (see Saatsi (2011a) for further discussion). One might wonder how much we gain from this, given that arguably Galilean idealisations do not present a serious challenge to realism to begin with. Although I already admitted (§2) that realists have a wealth of resources in responding to a challenge posed by Galilean idealisations, I think the realist can further gain from the conceptual framework advocated here. In particular, the framework allows us to shed further light on the modal aspects of idealised models, and how those aspects can feed into an account of an idealised model's empirical success. This framework affords us a better sense of a particular way in which an idealised model can latch onto reality so as to account for the model's empirical success.

Furthermore, there are reasons to think that the framework can also deal with (at least some) non-Galilean idealisations. The distinction between Galilean and non-Galilean idealisations need not be as deep as one might think. In relation to the much discussed infinite continuum idealisations in statistical physics, for example, we may construe the distinction in terms of how indispensable a given idealisation is to a model. On one side we have Galilean idealisations which are *controllable*, at least in principle, in the sense that we can we can replace our original model with a related, less idealised model that represents the system in question more truthfully (for example by including previously omitted forces). On the other side we have *uncontrollable*, non-Galilean idealisations that cannot be thus eliminated or reduced, even in principle, by a related, less idealised model. A paradigmatic example of such uncontrollable idealisation is the use of the thermodynamic limit in statistical physics of finite systems, where the number of particles n and the volume V of a system are taken to infinity while keeping n/V constant. This mathematical idealisation is uncontrollable since it cannot be replaced with a model that takes n to be some finite-but-large number (e.g. $\sim 10^{23}$), thereby representing better the finitude and atomicity of the actual system.

The sense of indispensability of such uncontrollable idealisations raises interesting questions, and it clearly in some sense demarcates these idealisations from the controllable cases. The uncontrollability in and of itself does not mean that these models cannot be viewed as inferentially veridical, however. What it means, rather, is that we are unable to construct models that are more veridical in these idealising

respects, so as to demonstrate *in that way* how the predictive success of the idealised model is robust under variation in the idealising assumptions. Models incorporating uncontrollable idealisations can still be inferentially veridical, however, in the sense that it can be a modal fact about the relevant laws of nature that they deductively yield, when combined with more veridical assumptions about the idealised features, the same or improved degree of empirical adequacy. *Our* (in) ability to demonstrate this—in principle or in practice—by de-idealising the original model need not necessarily be taken to indicate that such fact does not obtain.

There is a close analogy here with debates concerning mathematics' indispensability to science. Nominalists argue that regardless of our inability to nominalize our best theories we can maintain that it is the non-mathematical content of our theories that is responsible for the theories' empirical success, with mathematics playing a role only in representing non-mathematical facts and facilitating reasoning about it.[15] In a similar spirit I maintain that the indispensability of the uncontrollable infinite limits in statistical mechanics, for instance, can be indispensable only for representing and reasoning about systems with enormous but nevertheless finite numbers of components. It can still be a modal fact about the relevant micro-level laws of nature that they entail the same empirical results from veridical assumptions about the interacting micro-constituents.

But how, one may wonder, can this attitude be justified, if not by having good reasons to think that a model is de-idealisable, at least in principle? The answer is that one's understanding of the workings of an uncontrollable idealisation can involve much else besides the assumptions that go into a particular non-Galilean model. That is, the full set of theoretical resources that can come to bear on justifying one's belief in such modal fact about the laws—justifying the inferential veridicality of the idealised model—goes well beyond the modelling assumptions. In the full theoretical context of such models we can arguably explain, by reference to relevant facts about finite systems, why an infinite mathematical idealisation is empirically adequate to the degree it is, notwithstanding its indispensability. This broader theoretical contexts has been extensively discussed in the recent literature (see e.g. Butterfield 2011a, b; Menon and Callender 2013; Norton 2012). It is through such theoretical accounts of a given uncontrollable idealisation that we get a handle on the sense in which the model 'gets right' some critical less specific features of large-enough systems. These are the features that the model shares with large finite systems, features that in conjunction with the relevant laws entail the right predictions (to a sufficient degree of approximation).

The details of these 'reductionist accounts' of the continuum limit in statistical physics remain to be discussed further in the context of my conceptual framework. I have to leave this for further work, and move on to conclude the paper with brief remarks on explanation. Throughout the paper I have focused on the predictive success of idealised models, largely bracketing the role of idealisations in successful scientific explanations. The explanatory dimension also matters to the realist, of

[15]See e.g. Melia (2000) and Saatsi (forthcoming).

course, given the role of inference to the best explanation in many realist gambits, for example. (It is worth noting that Batterman's much discussed work on uncontrollable idealisations almost exclusively concern their explanatory indispensability.) It is impossible for me to do justice to this rather large topic here, but let me just note the importance of considering models' explanatory successes quite separately from their predictive successes. The distinction between predictive and explanatory success was perhaps only of minor consequence back in the day of the DN-model of explanation. But in the contemporary context, largely ruled by different modal accounts of explanation, the conceptual difference between prediction and explanation matters a great deal to the way realists should apportion their epistemological commitments in relation to scientifically successful theories and models. Different issues come to the fore in accounting for the explanatory role played by the falsehoods that constitute idealisations. The indispensability of idealisations for explanations, for example, raises issues for the realist that are closely related, or analogous to the issues raised by the arguably indispensable role that mathematics plays in scientific explanations. I have argued elsewhere that the realist should consider the latter issues in close contact with well-formed views about the nature explanation (Saatsi forthcoming). I believe the same holds for the former issues as well.

Acknowledgments A version of this paper was presented at a workshop on Models and Inferences in Science in Rome. Thanks to the workshop audience, as well as James Fraser, Steven French, and especially Arnon Levy, for helpful comments.

References

Batterman, R.W.: Critical phenomena and breaking drops: Infinite idealizations in physics. Stud. Hist. Philos. Sci. Part B Stud. Hist. Philos. Mod. Phys. **36**(2), 225–244 (2005)

Bueno, O., French, S.: Can Mathematics Explain Physical Phenomena?, Br. J. Philos. Sci. **63**(1), 85–113 (2012)

Butterfield, J.: Emergence, reduction and supervenience: A varied landscape. Found. Phys. **41**(6), 920–959 (2011a)

Butterfield, J.: Less is different: Emergence and reduction reconciled. Found. Phys. **41**(6), 1065–1135 (2011b)

Da Costa, N., French, S.: Science and partial truth: A unitary approach to models and scientific reasoning. Oxford University Press, USA (2003)

Giere, R.: Explaining science: A cognitive approach. University of Chicago Press, Chicago (1988)

Lewis, D.: On the plurality of Worlds. Blackwell, Oxford (1986)

McMullin, E.: Galilean idealization. Stud. Hist. Philos. Sci. **16**, 247–273 (1985)

Melia, J.: Weaseling away the indispensability argument. Mind **109**, 455–479 (2000)

Menon, T., Callender, C.: Ch-Ch-Changes philosophical questions raised by phase transitions. In: Robert, B. (ed.) The Oxford Handbook of Philosophy of Physics, pp. 189. OUP, USA (2013)

Niiniluoto, I.: Idealization, counterfactuals, and truthlikeness. In Jerzy, B., Andrzej, K., Theo, A. F. K., Krzysztof, L., Katarzyna, P., Piotr, P. (eds.) The Courage of Doing Philosophy: Essays Dedicated to Leszek Nowak, pp. 103–122. Rodopi (2007)

Norton, J.: Approximation and idealization: Why the difference matters. Philos. Sci. **79**(2), 207–232 (2012)

Odenbaugh, J.: True lies: Realism, robustness and models. Philos. Sci. **78**(5), 1177–1188 (2011)

Saatsi, J.: Idealized models as inferentially veridical representations: A conceptual framework. In: Humphreys, P., Imbert, C. (eds.) Representations, Models, and Simulations, pp. 234–249. Routledge, London (2011a)

Saatsi, J.: The enhanced indispensability argument: Representational versus explanatory role of mathematics in science. Br. J. Philos. Sci. **62**(1), 143–154 (2011b)

Saatsi, J. (forthcoming). On the 'Indispensable Explanatory Role' of Mathematics. Mind

Saatsi, J., Vickers, P.: Miraculous success? Inconsistency and untruth in Kirchhoff's diffraction theory. Br. J. Philos. Sci. **62**(1), 29–46 (2011)

Saatsi, J.T.: Reconsidering the Fresnel-Maxwell case study. Stud. Hist. Philos. Sci. **36**, 509–538 (2005)

Sorensen, R.: Veridical idealizations. In: Frappier, M., Meynell, L., Brown, J. R. (eds.) Thought Experiments in Philosophy, Science, and the Arts, pp. 30–50. Routledge (2012)

Strevens, M.: Depth: An Account of Scientific Explanation. Harvard University Press, Harvard (2008)

Toon, A.: Models as Make-Believe: Imagination, Fiction, and Scientific Representation. Palgrave Macmillan, New York (2012)

Modelling Non-empirical Confirmation

Richard Dawid

Abstract The paper provides a presentation and motivation of the concept of non-empirical theory confirmation. Non-empirical theory confirmation is argued to play an important role in the scientific process that has not been adequately acknowledged so far. Its formalization within a Bayesian framework demonstrates that non-empirical confirmation does have the essential structural characteristics of theory confirmation.

1 Introduction

The canonical view of the scientific process understands theory confirmation in terms of a direct confrontation of a theory's predictions with empirical data. A scientific theory is expected to make testable empirical predictions. If the relevant collected data agrees with those predictions, the data confirms the theory. If the data disagrees with the predictions, the theory gets disconfirmed.

One may view this understanding in terms of a technical definition of theory confirmation, which would render it immune against criticism. It may be argued, however, that the concept of confirmation should account for the scientists' actual reasons for taking a theory to be well-established as a viable description of a given aspect of the observed world. If that aim is endorsed, one may question a given understanding of theory confirmation by comparing it with the scientists' actual attitude towards their theories.

The latter view is the point of departure chosen in the present article. It is assumed that the concepts deployed by the philosophy of science for modelling

R. Dawid (✉)
LMU Munich—Munich Center for Mathematical Philosophy, Munich, Germany
e-mail: richard.dawid@univie.ac.at
URL: http://homepage.univie.ac.at/richard.dawid/

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_11

scientific reasoning should offer a characterization of the actual structure of scientific reasoning—and should be measured by that standard. On that account, however, a closer look at actual science throws the adequacy of the canonical understanding of theory confirmation into doubt. In many scientific fields, confirmation in the canonical sense described above is not the only basis for an assessment of a theory's status. Three interrelated issues arise, which render a full focus on empirical confirmation insufficient. They shall be briefly sketched in the following.

1: In historical fields of research, scientists often face a conjunction of two problems. First, the general character of scientific hypotheses in those fields often makes it difficult to extract specific and quantitative predictions from them. Second, and maybe even more troubling, those scientific fields often deal with empirical situations where most of the empirical record has been irretrievably lost to natural decay or destruction during the periods that lie between that events under investigation and the time of inquiry. Moreover, even of the data that would be available in principle, it is often only possible to collect a haphazard and arbitrary subset.[1] Anthropologists searching for early human traces, to give one example, cannot search specifically for the missing link they are most interested in but must be content with whatever new material their excavations provide. The two described conditions in conjunction create a situation where empirical confirmation remains patchy and, on its own, does not provide a stable foundation for assessing the probability that a theory is trustworthy. External characteristics of the theory and the research field therefore play an important role in that assessment.

More specifically, if various conflicting hypotheses aim at explaining the same available data, abductive forms of reasoning are deployed, which depend on understanding whether or not one of the theories seems substantially more plausible than the others. One important issue that must be addressed in such cases is the question whether and if so on what grounds it makes sense to assume that those theories that have been developed cover the spectrum of possible plausible theories on the issue. Only if that is the case does it make sense to trust the most plausible of the known theories. Trust in a theory thus is instilled based on a combination of assessments of the spectrum of known alternatives and some induced understanding of the presumptive spectrum of unconceived alternatives.

2: A similar issue arises in the case of micro-physical theories that conjecture the existence of unobservable physical objects like atoms, quarks or, to mention an important recent example, the Higgs particle. In those cases, announcing a discovery amounts to endorsing all empirical implications of the discovered

[1]For an instructive philosophical perspective on historical sciences, see Turner (2007).

object, whether or not they have been empirically tested yet. A discovery therefore has profound consequences in high energy physics. Once a particle has been discovered in one experiment, background calculations in all future experiments factor in all empirical implications the particle has within the established theoretical framework. The question is, however, on what basis scientists can be so confident that no unconceived alternative could account for the data collected without having the same further empirical implications as the known theory. The answer is that scientists cannot make that assessment without relying on observations about the overall research process. They need to make an assessment as to whether or not an alternative seems likely to show up based on their understanding of the overall conceptual context and whether assessments of that kind have turned out reliable in the past. In other words, the declaration of a discovery of a new object in microphysics relies on considerations very similar to those which lead scientists towards endorsing a theory in palaeontology or other non-formalized historical sciences.

3: Finally, since the 1980s high energy physicists and cosmologists have shown an increasing readiness to invest a high degree of trust in empirically unconfirmed theories. Theories like string theory or cosmic inflation are taken by many as important steps towards a deeper understanding of nature even though those theories so far have no (in the case of string theory) or only inconclusive (in the case of inflation) empirical confirmation. Once again it turns out that the reasons responsible for that trust are of a very similar kind as those at play in the previously discussed contexts.

Unlike in the previously discussed cases, the extent to which exponents of empirically unconfirmed theories in fundamental physics consider their theory well-established has led to a highly controversial debate on the scientific legitimacy of the involved strategies of theory assessment. In this light, the case of scientific trust in empirically unconfirmed theories turns the question of an adequate understanding of the concept of confirmation from a mainly philosophical issue into a question of high significance for the further evolution of fundamental physics.

All three discussed scientific contexts suggest that a perspective that focusses entirely on the agreement between a theory's predictions and empirical data is insufficient for acquiring an adequate understanding of the reasons why scientists trust a scientific theory.

In the following, I will present a widened concept of theory confirmation that, as I will argue, comes closer to that goal.

Two basic guidelines will determine the layout of the presented approach. On the one hand, as already pointed out above, the discussion will be guided by the idea that the concept of confirmation should provide a basis for understanding the degree of trust scientists have in a theory. On the other hand, however, the empirical character of science, that is the connection between confirmation and observation, shall not be abandoned. So, while the approach to be presented is non-canonical, it will be argued to remain true to core principles of scientificality.

2 The Setup

2.1 What Is Non-empirical Confirmation?

The canonical view as it is presented in accounts from classical hypothetico-deductivism to most readings of Bayesian confirmation (see e.g. Bovens and Hartmann 2003; Howson and Urbach 2006) constrains confirmation to observations within the theory's intended domain. Only the agreement between a theory's predictions and empirical data constitutes confirmation of that theory. We will call this form of confirmation "empirical confirmation" because it is based on empirical testing of the theory's predictions. Our question will be: which kinds of consideration beyond the limits of empirical confirmation may in principle be understood as contributions to theory confirmation? More specifically, we'll search for a form of "non-empirical" theory confirmation that can account for those considerations that have been argued above to be crucial in a number of contexts for instilling trust in a scientific theory.[2] At the same time, however, we want to retain the crucial role of observation and stay as close as possible to the mechanism of empirical confirmation. Confirmation should remain true to the basic principles of scientific reasoning.

In order to guarantee a grounding of confirmation in observation, we introduce elements of empiricist reasoning at two distinct levels. First, we understand trust in a theory in terms of the theory's empirical predictions rather than in terms of truth. If a scientist trusts a theory, she believes that the theory's predictions in its characteristic regime, if tested, will get empirically confirmed. If a theory's predictions in its characteristic regime are indeed in agreement with all possible data, the theory shall be called empirically viable. Non-empirical confirmation thus amounts to an increase of trust in the theory's empirical viability. Note that this understanding of confirmation in a certain respect stays closer to an empiricist understanding than concepts of empirical confirmation that are based on truth probability. By avoiding reference to truth, we block the possibility that theories which have no empirical implications get non-empirically confirmed. Trust in theories that have no empirical implications is trivial on our account and cannot be increased by any means. Therefore, confirmation of non-predictive theories cannot occur.

Second, we will require that confirmation be based on some observations about the world beyond the theory and its endorser. The mere fact that a consideration contributes to a person's subjective belief in a theory's viability does not justify calling that consideration non-empirical confirmation on our account. For example, the fact that some scientists trust elegant theories does not imply that a theory's elegance constitutes non-empirical confirmation.

[2]The concept was first laid out in Dawid (2006) and then further developed in Dawid (2013). A Bayesian formalization of one argument of non-empirical confirmation was given in Dawid et al. (2015).

Being based on observations about the world is a fairly vague requirement, however. Which kind of relation between observation and confirmation do we require? One might follow various strategies in this regard. One plausible guideline, which we will follow, is a structural similarity between empirical and non-empirical confirmation.

We will introduce the following fairly specific definition of non-empirical confirmation. Non-empirical confirmation is based on observations about the research context of the theory to be confirmed. Those observations lie within the intended domain of a meta-level hypothesis about the research process and, in an informal way, can be understood to provide empirical confirmation of that meta-level hypothesis. The meta-level hypothesis, in turn, is positively correlated with the probability of the truth or viability of the scientific theory under scrutiny.

This may seem like a fairly complicated and arbitrary construction at first glance. However, it has a number of considerable merits. Most significantly, non-empirical confirmation of the suggested kind turns out to work as a reconstruction of the most conspicuous lines of reasoning that do generate trust in a scientific theory beyond the limits of empirical confirmation.

Second, non-empirical confirmation in the suggested sense can be understood in terms of an extension of the basis of observational evidence for a theory. The mechanisms of connecting observations to the overall conceptual framework remain the same as in the case of empirical confirmation. Confirmation is still based on comparing predictions with observations, but that comparison may play out at the meta-level of analysing the research process within which the theory is embedded rather than at the ground level of the theory's subject matter.

Third, and directly related to point 2, non-empirical confirmation of the described kind resembles empirical confirmation in being symmetric between confirmation and dis-confirmation. Observations at the meta-level may equally support and speak against a theory's viability, depending on whether or not they agree with the predictions of the meta-level hypothesis. The correlation between observational input and confirmation/dis-confirmation thus works along very similar lines as in empirical confirmation.

2.2 Towards a Formalized Model

Conformation today is mostly understood in Bayesian terms. In this light, we will analyse the nature of non-empirical confirmation from a Bayesian perspective. It will turn out that a probabilistic approach is particularly suitable for characterizing the way non-empirical confirmation works.

In Bayesian terms, an increase of trust in a theory's viability is expressed as an increase of the subjective probability that the theory is viable. As already discussed, our use of probabilities of viability constitutes a deviation from canonical Bayesian epistemology, which is based on truth probabilities.

We introduce the proposition T that a theory H is empirically viable (consistent with all empirical data) within a given context. Let us first consider the case of empirical confirmation. We take H to be confirmed by empirical data E iff $P(\mathrm{T}|\mathrm{E}) > P(\mathrm{T})$, that is if the subjective probability of the viability of H is increased by E. If E lies within the extended domain of H, one can deduce a probability of E from H and a set of initial conditions specified based on other observations. A high probability of E then justifies

$$P(E|T) > P(E) \tag{1}$$

which implies that E confirms H due to Bayes' formula

$$\frac{P(T|E)}{P(T)} = \frac{P(E|T)}{P(E)} \tag{2}$$

Now our goal is to replace E by some observations F^X that are not in the intended domain of H. In other words, H in conjunction with knowledge about initial conditions for the system described by H does not provide any information on the probability of F^X. Nevertheless F^X should imply $P(\mathrm{T}|\mathrm{F}^X) > P(\mathrm{T})$.

Further, we want this probability increase to be induced via a new hypothesis Y that lives at the meta-level of theory assessment and is positively correlated with both F and T. Moreover, F^X should be in the intended domain of Y, that is, implications for F^X can be extracted from hypothesis Y.

In principle, one might try to find a specific variable Y^X for each type of non-empirical observation F^X. However, we shall pursue a different strategy and specify one Y that will be tested by various forms of F^X. This strategy has two advantages. First, it turns out to work well with respect to the three most conspicuous candidates for non-empirical theory confirmation to be found in science. And second, it allows for a more coherent overall understanding of the way the arguments of non-empirical confirmation mutually support each other.

So what would be a plausible candidate for Y? It is helpful to think about this question by looking at the most straightforward candidate for an argument of non-empirical theory confirmation: the no alternatives argument (see Dawid et al. 2015). Let us, for the time being, continue the analysis within the framework of this specific argument. Later, we will return to a more general perspective.

3 The No Alternatives Argument

Let us assume that we make the following observation F^A: scientists have looked intensely and for a considerable time for alternatives to a known theory H that can solve a given scientific problem but haven't found any. This observation may be taken by us as an indication that the theory they have is probably viable.

Clearly, this kind of reasoning plays an important role in generating trust in some empirically unconfirmed or insufficiently confirmed theories in science. As mentioned above, a specific reconstruction of a phenomenon or object in anthropology or other historic sciences gains credibility if the case can be made that no other plausible reconstruction has been found. Most high energy physicists believed in the existence of a Higgs particle even before its discovery in 2012 because no satisfactory explanation of the mass spectrum of elementary particles that did not rely on some kind of Higgs particle had been found.

We call this kind of reasoning the no alternatives argument (NAA) (Dawid et al. 2015). In the following, we give a Bayesian reconstruction of NAA. In the case of NAA, we can easily identify the most natural candidate for Y: to the extent F^A increases the probability of the viability of H, it arguably does so by supporting the hypothesis that there in fact are no or very few possible scientific alternatives to H. NAA thus involves an inference from an observation F_A on the alternatives discovered to a statement Y on the actual number of possible alternatives. Y thus is a statement on the limitations on the number of possible scientific theories on a subject. In order to make sense of this, we need to specify the framework more clearly. Let us assume a theory H that is built to account for empirical data $\mathcal{D}$. We now assume that there exists a specific but unknown number i of possible scientific theories (i.e. theories that satisfy a set of scientificality constraints $\mathcal{C}$) which are compatible with the existing data $\mathcal{D}$ and give distinguishable predictions for the outcome of some relevant set $\mathcal{E}$ of future experiments.

Here, $\mathcal{D}$ specifies the empirical status quo. Possible scientific theories on the subject must be consistent with the relevant available data $\mathcal{D}$. In the most straightforward cases, H can be shown either to predict data $\mathcal{D}$ or at any rate to be consistent with it. There are also more difficult cases (like e.g. string theory) where consistency of H with data $\mathcal{D}$ has not been established but is considered plausible. Obviously, a situation of the latter type generates a comparably lower prior probability $P(T)$. Still non-empirical confirmation can work on that basis as well.

$\mathcal{C}$ specifies what counts as a scientific theory. Only those theories that meet the scientificality-conditions $\mathcal{C}$ count as possible theories. Scientificality-conditions are themselves volatile to a given degree an may change in time. Note, however, that our formal argument does not rely on a precise specification of the scientificality-conditions. All we need is the assumption that scientists apply a set of scientificality-conditions that contains a viable core that can be satisfied by theories that are empirically viable with respect to the future experiments $\mathcal{E}$.

Having introduced a framework for scientific theory building, we still need to specify a way of individuating theories. We need to decide up to which point we still speak of one theory and when we start talking about two different theories.

Generally speaking, we individuate theories by their predictive implications with respect to future experiments $\mathcal{E}$. Theories which give the same predictions (or the same range of predictions under variation of their parameter values) count as one theory. The reason for choosing this approach is that we are mainly interested in

empirical predictions. Trust in a theory, from our perspective, is justified if the theory ends up making correct predictions. Since we only talk about empirical viability and not about truth, questions related to the spectrum of empirically equivalent theories lie beyond the scope of our analysis.

The specific form of criteria for theory individuation depends on the kind of predictions one is interested in. Therefore, we don't prescribe those criteria in detail. Scientists implicitly select them in dependence on the level at which they trust their theory's predictions. Let us explicate this by looking at the example of the empirical testing of the Higgs model. The Higgs model is a theoretical concept that can explain why elementary particles have masses. It predicts the existence of a scalar particle with a mass that lies within a certain range of possible values. Physicsts had a high degree of trust in the existence of the Higgs particle already long before the particle's discovery in 2012. Let us now assume that, before 2012, some physicist wanted to predict that the Higgs-particle existed and had precisely mass M_1. This 'theory', let us call it H_1, would have been distinct from any other exemplification of the Higgs model that predicted a different mass for the Higgs particle. In order to count the alternatives to H_1 one would have had to count each of these variations as an individual 'theory' H_n and thus would have got an infinite number of possible alternatives to H_1. Given that physics before 2012 did not offer arguments for the precise Higgs mass, it would have been clear that one could not trust H_1 or its predictions. Individuating theories based on specific Higgs masses thus would have been an inadequate basis for deploying NAA with respect to the Higgs hypothesis.

Since there was no basis for predicting the precise Higgs mass before 2012, physicists were most interested in the question as to whether the Higgs particle exists at all without specifying its precise mass. They were interested in the general viability of the Higgs hypothesis as a theoretical mechanism that could explain particle masses and implied the existence of at least one scalar particle—and they were quite confident about the viability of the Higgs mechanism even in the absence of empirical data. When analysing this situation, an assessment of possible alternatives to the Higgs hypothesis must not count different mass values as different theories. Even the specification of the Higgs model beyond its core structure (by introducing additional scalars, a constituent structure of the Higgs particle, etc.) would not have counted as a different theory at this level. Only substantially different approaches to mass generation which did not rely on a scalar field would have counted as alternatives to the Higgs hypothesis. The fact that one had not found any convincing alternatives at this level gave reason to trust in the viability of the Higgs hypothesis even in the absence of empirical confirmation. The level of theory individuation used in NAA had to correspond to this line of reasoning.

Having thus clarified the framework for specifying the number i of possible alternatives to theory H, we can now proceed to the proof that NAA amounts to confirmation of H in Bayesian terms based on a set of very plausible assumptions (Fig. 1).

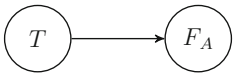


Fig. 1 The Bayesian Network representation of the two-propositions scenario

4 Formalizing the No Alternatives Argument

We introduce the binary propositional variables T and F_A, already encountered in the previous section.[3] T takes the values

T The hypothesis H is empirically viable.
¬T The hypothesis H is not empirically viable.

and F_A takes the values

F_A The scientific community has not yet found an alternative to H that fulfills $\mathcal{C}$, explains $\mathcal{D}$ and predicts the outcomes of $\mathcal{E}$.
$\neg F_A$ The scientific community has found an alternative to H that fulfills $\mathcal{C}$, explains $\mathcal{D}$ and predicts the outcomes of $\mathcal{E}$.

We would now like to explore under which conditions F_A confirms H, that is, when

$$P(T|F_A) > P(T). \tag{3}$$

We then introduce variable Y that mediates the connection between T and F_A. In the previous section, we characterized Y in general terms as a statement about limitations to the number of possible alternatives to theory H. In our formalization, we are more specific. Y has values in the natural numbers, and Y_k corresponds to the proposition that there are exactly k hypotheses that fulfil $\mathcal{C}$, explain $\mathcal{D}$ and predict the outcomes of $\mathcal{E}$.

The value of F_A—that scientists find/do not find an alternative to H—does not only depend on the number of available alternatives, but also on the relation between the difficulty of the problem and the capabilities of the scientists. Call the variable that captures this factor S, and let it take values in the natural numbers, with $S_j := \{S = j\}$ and $d_j := P(S_j)$. The higher the values of S, the more difficult the problem is to solve for the scientists.[4] It is clear that S has no direct influence on Y and T (or vice versa), but that it matters for F_A and that this influence has to be represented in our Bayesian Network.

We now list five plausible assumptions that we need for showing the validity of the No Alternatives Argument.

[3]The presentation of this section is largely taken from Dawid et al. (2015).

[4]For the purpose of our argument, it is not necessary to assign a precise operational meaning to the various levels of S. It is sufficient that they satisfy a natural monotonicity assumption with regard to their impact on F_A—see condition **A3**.

A1. The variable T is conditionally independent of F_A given Y:

$$T \perp\!\!\!\perp F_A | Y \tag{4}$$

Hence, learning that the scientific community has not yet found an alternative to H does not alter our belief in the empirical adequacy of H if we already know that there are exactly k viable alternatives to H.

A2. The variable D is (unconditionally) independent of Y:

$$D \perp\!\!\!\perp Y \tag{5}$$

Recall that D represents the aggregate of those context-sensitive factors that affect whether scientists find an alternative to H, but that are not related to the number of suitable alternatives. In other words, D and Y are orthogonal to each other by construction.

These are our most important assumptions, and we consider them to be eminently sensible. Figure 2 shows the corresponding Bayesian Network. To complete it, we have to specify the prior distribution over D and Y and the conditional distributions over F_A and T, given the values of their parents. This is done in the following three assumptions.

A3. The conditional probabilities

$$f_{kj} := P(\mathrm{F_A} | \mathrm{Y_k}, \mathrm{D_j}) \tag{6}$$

are non-increasing in k for all $j \in \mathbb{N}$ and non-decreasing in j for all $k \in \mathbb{N}$. The (weak) monotonicity in the first argument reflects the intuition that for fixed difficulty of a problem, a higher number of alternatives does not decrease the likelihood of finding an alternative to H. The (weak) monotonicity in the second argument reflects the intuition that increasing difficulty of a problem does not increase the likelihood of finding an alternative to H, provided that the number of alternatives to H is fixed.

A4. The conditional probabilities

$$t_k := P(\mathrm{T} | \mathrm{Y_k}) \tag{7}$$

are non-increasing in k.

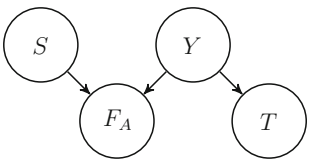


Fig. 2 The Bayesian Network representation of the four-propositions scenario

This assumption reflects the intuition that an increase in the number of alternative theories does not make it more likely that scientists have already identified an empirically adequate theory.

A5. There is at least one pair (i, k) with $i < k$ for which (i) $y_i y_k > 0$ where $y_k := P(\mathrm{Y_k})$, (ii) $f_{ij} > f_{kj}$ for some $j \in \mathbb{N}$, and (iii) $t_i > t_k$.
In particular, this assumption implies that $y_k < 1$ for all $k \in \mathbb{N}$ because otherwise, a pair satisfying (i) could not be found.

With these five assumptions, we can show that (For a proof, see Dawid et al. 2015):

Theorem 1 *If Y takes values in the natural numbers* $\mathbb{N}$ *and assumptions* **A1** *to* **A5** *hold, then* $\mathrm{F_A}$ *confirms* T, *that is*, $P(\mathrm{T}|F_\mathrm{A}) > P(\mathrm{T})$.

$\mathrm{F_A}$ thus confirms the empirical viability of H under rather weak and plausible assumptions.

5 The Meta-Inductive Argument

So NAA formally constitutes confirmation. The question remains, however, how significant that confirmation is. The problem is that we have two possible explanations of F_A. F_A may be explained by the fact that there are no or very few possible alternatives to H. However, it might also be explained by a statement of type S: scientists are not clever enough to find the complicated alternatives that are possible. F_A cannot distinguish between those two kinds of explanation. If it is our prior assessment that an explanation of type S is far more likely to apply than an explanation of type Y, even the most powerful observation F_A could not alter this assessment. Therefore, if we start with very low priors for low number Y_ks and high priors for the hypothesis that scientists are not clever enough to find most of the alternatives, F_A won't strongly increase probabilities for low number Y_ks and therefore won't provide significant confirmation of H.

In order to turn NAA into a significant argument, we therefore need a second line of reasoning that allows us to distinguish between S and Y and, on that basis, can establish considerable probabilities of low number Y_ks which can then serve as a basis for significant confirmation of H by NAA.

This brings us to the second argument of non-empirical confirmation, the meta-inductive argument (MIA). The meta inductive argument is based on the observation F_M that those theories in the research field that satisfy a given set of conditions K (note that these are not the scientificality conditions C but may be considerably more restrictive) have shown a tendency of being viable in the past.

A meta-inductive step leads directly from F_M to inferring a high probability $P(T|F_M)$. However, in order to use MIA as support for NAA, it is helpful once again to use the statements Y as an intermediary. In order to do so, we have to assume a stronger connection between empirical viability and the number of alternatives. The simplest and most straightforward assumption would be that the theory found by

the scientists is a random pick from the set of empirically distinguishable possible alternatives. This means that a theory's chances of being viable is $P(T) = 1/i$. Based on this model one can understand our subjective probability $P(T)$ in terms of our assessment of the probabilities of numbers of alternatives. For the simple model introduced above we get

$$P(T) = \sum_k P(Y_k)P(T|Y_k) = \sum_k \frac{1}{k} P(Y_k). \tag{8}$$

On that basis, if one observes a certain success rate of theories in a research field that satisfy conditions K, a frequentist analysis enforces substantial probabilities for Y_ks with low k. To give an example, let us assume that we observe a success rate of 50 % of theories that satisfy K. A simple calculation shows that, based on our model and frequentist data analysis, we must attribute a probability of 1/3 or higher to the hypothesis ($k = 1 \vee k = 2$). MIA therefore generates assessments of $P(Y_k)$ which can then serve as priors in NAA.

MIA thus strengthens explanation Y of F_A and weakens explanation S correspondingly. If scientists were so successful in finding viable theories in the past, it seems less plausible to assert that they are not clever enough for doing the same this time. Therefore, MIA can turn NAA into a method of significant confirmation.

One important worry may arise even if MIA and NAA look strong: it is not a priori clear whether the empirically unconfirmed theory that is evaluated is sufficiently similar in relevant respects to earlier successful theories to justify meta-inductive inference from the viability of those earlier theories to the viability of the present one.

Now it may happen that the theory under evaluation is so closely related and the problems addressed are so similar to earlier cases that there just seems no plausible basis for that worry. The Higgs hypothesis is an example of such a scenario. It is so deeply immersed in standard model physics that it would be quite implausible to argue that physicists understood the other problems raised with respect to the standard model sufficiently well but were not up to the complexity of the Higgs case.

In a similar vein, certain principles of reconstructing species from scarce excavated evidence may be applicable in many fairly similar individual cases. If such a strategy has proved successful in a number of cases, this may be sufficient for trusting the reliability of the method in similar cases in the future, provided the method offers the only known plausible reconstruction of the given species.

In cases like those referred to above, NAA and MIA in conjunction can be sufficient for generating a high degree of trust in a theory. The trust in the Higgs mechanism was indeed generated largely by those two arguments: the understanding that no convincing alternative to the Higgs mechanism was forthcoming and the observation that standard model physics had turned out predictively extremely successful whenever it had been tested.

6 Unexpected Explanatory Interconnections

There are other cases, however, where differences between previous theories and the new one with respect to the nature or the complexity of the core problems are very significant. String physics is a particularly good example of such a situation. Though string theory stands in the tradition of previous high energy physics, it clearly is a far more complex and difficult theory than any of its forebears. Thus it may easily be argued that, while scientists were clever enough to understand the spectrum of possibilities in the case of standard model physics, they are not up to the task with respect to string physics.

In cases like this, a third argument is required in order to turn NAA + MIA into a convincing line of reasoning. Arguably, the most effective argument of this kind is the argument of unexpected explanatory interconnections (UEA). The observation F_U on which this argument is based is the following. Theory H was developed in order to solve a specific problem. Once H was developed, physicists found out that H also provides explanations with respect to a range of problems which to solve was not the initial aim of developing the theory.

The argument is structurally comparable to the argument of novel empirical confirmation: a theory that was developed in order to account for a given set of empirical data correctly reproduces data that had not entered the theory's construction process. UEA is the non-empirical "cousin" of novel confirmation: instead of successful novel empirical predictions, the theory provides unexpected explanatory interconnections that do not translate into successful empirical predictions.

The most prominent example of UEA in the context of string theory is the microphysical derivation of the black hole entropy area law in special cases of black holes. String theory was not developed for providing this derivation. More than two decades after string theory was proposed as a theory of all interactions, Strominger and Vafa (2006) succeeded in providing it.

UEA fits well into the category of non-empirical confirmation because it can be read as an argument for limitations to underdetermination just like NAA and MIA. The line of reasoning in the case of UEA is the following. Let us assume a set of n seemingly unrelated scientific problems in a research field. Let us further assume that there is a number i of possible alternative solutions to one of those problems. If the number of possible solutions to a specific problem is typically much higher than n, we have no reason to expect that a random solution to one problem will solve other problems as well. If we assume, however, that i is typically substantially smaller than n, we may plausibly assume that consistent strategies for solving one individual problem will typically be applicable to a number of problems. The reason for this is that we know that there is one theory, the true theory, that does solve all n problems. Therefore, in the extreme case that there is only one consistent solution to the problem we look at, all problems must be solved by that theory. Inversely, the observation that the theory that was developed for solving the given problem turns out to answer a number of other open questions as well can be taken

as an indicator that there probably are very few possible different solutions to the given problem. From that consideration, once again, there follows a reasonably high probability that the given theory is viable.

UEA is of particular importance in contemporary fundamental physics, where theory building gets extremely difficult and UEA can provide a check as to whether or not physicists are capable of dealing with the overall set of problems they face in a way that goes beyond limited puzzle solving with respect to individual problems.

7 Conclusion

What is the status on non-empirical theory confirmation? As already emphasised in the introduction, non-empirical confirmation is an extension of empirical confirmation with a widened arsenal of conceptual tools but similar basic patterns of reasoning. It is secondary to empirical confirmation for two reasons. First, non-empirical confirmation is understood as a tool for establishing a theory's viability. Viability however, as defined in Sect. 2, is based on the theory's agreement with empirical data. Therefore, the perspective of eventual empirical testing is always in the background and, once conclusive empirical testing can be achieved, will in all cases make a stronger case for the theory's viability than non-empirical confirmation ever could. Second, the significance of non-empirical confirmation crucially relies on MIA. MIA, however, as described in Sect. 5, is based on empirical confirmation somewhere else in the research field. Non-empirical confirmation therefore can only work properly as long as empirical confirmation can be achieved somewhere in the research field.

Non-empirical confirmation is closely linked to a probabilistic view on confirmation. To a philosopher who denies that confirmation has anything to do with attributing a probability to a theory's viability or truth, non-empirical confirmation will look empty. On the other hand, to a philosopher who acknowledges a probabilistic basis of confirmation, it seems difficult to deny that non-empirical confirmation exists. From that perspective, the core question becomes how significant non-empirical confirmation can be. This paper offered some indications as to how a formal argument in favour of the significance non-empirical confirmation could be developed.

It has to be emphasised that a general argument for the legitimacy of non-empirical confirmation by no means implies that each individual deployment of non-empirical confirmation is convincing. There are cases in science, some of which have been mentioned in this paper, where the actual strength and influence of non-empirical arguments for a theory's viability is indeed striking. There are many others where understanding the strengths and weaknesses of the use of non-empirical confirmation requires careful analysis. I suggest that a probabilistic account of non-empirical confirmation provides a promising framework for carrying out that kind of analysis in a fruitful way.

References

Bovens, L., Hartmann, S.: Bayesian epistemology. Oxford University Press, Oxford (2003)
Dawid, R.: Underdetermination and theory succession from the perspective of string theory. Philos. Sci. **73**, 298–322 (2006)
Dawid, R.: String theory and the scientific method. Cambridge University Press, Cambridge (2013)
Dawid, R., Hartmann, S., Sprenger, J.: The no alternatives argument. Brit. J. Philos. Sci. **66**(1), 213–234 (2015)
Howson, C., Urbach, P.: Scientific reasoning: the Bayesian approach, 3rd edn. Open Court, La Salle (2006)
Strominger, A., Vafa, C.: Microscopic origin of the Bekenstein Hawking entropy. Phyiscs lett. **B379**, 99–104 (2006)
Turner, D.: Making prehistory: historical science and the scientific realism debate. Cambridge University Press, Cambridge (2007)

Mathematics as an Empirical Phenomenon, Subject to Modeling

Reuben Hersh

Abstract Among the universal attributes of homo sapiens, several have become established as special fields of study—language, art and music, religion, political economy. But mathematics, another universal attribute of our species, is still modeled separately by logicians, historians, neuroscientists, and others. Could it be integrated into "mathematics studies," a coherent, many-faceted branch of empirical science? Could philosophers facilitate such a unification? Some philosophers of mathematics identify themselves with "positions" on the nature of mathematics. Those "positions" could more productively serve as models of mathematics.

Modeling, the topic of this meeting, is a central feature of contemporary empirical science. There is mathematical modeling, there is computer modeling, and there is statistical modeling, which is half way between. We may recall older models: plaster models of mathematical surfaces, stick-and-ball models of molecules, and the model airplanes that used to be so popular, but now have been promoted into drones.

Today the scholarly or scientific study of any phenomenon, whether physical, biological, or social, implicitly or explicitly uses a model of that phenomenon. A physicist studying heat conduction, for example, may model heat conduction as a fluid flow, or as propagation of kinetic energy of molecules, or as a relativistic or quantum mechanical action. Different models serve different purposes. Setting up a model involves focusing on features of the phenomenon that are compatible with the methodology being proposed, and neglecting features that are not compatible with it. A mathematical model in applied science explicitly refrains from attempting to be a complete picture of the phenomenon being modeled.

Mathematical modeling is the modern version of both applied mathematics and theoretical physics. In earlier times, one proposed not a model but a theory. By talking today of a model rather than a theory, one acknowledges that the way one studies the phenomenon is not unique; it could also be studied other ways. One's

R. Hersh (✉)
1000 Camino Rancheros, Santa Fe, NM 87505, USA
e-mail: rhersh@gmail.com

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_12

model need not claim to be unique or final. It merits consideration if it provides an insight that isn't better provided by some other model.

It is disorienting to think of mathematics as the thing being modeled, because much of mathematics, starting with elementary arithmetic, already *is* a model of a physical action. Arithmetic, for instance, models the human action of counting.

Philosophy of mathematics, when studying the "positions" of formalism, constructivism, Platonism, and so on, is studying models of mathematics, which is in large part a model. It studies second-order models! (Other critical fields like literary and art criticism are also studying models of models.) Being a study of second-order models, philosophy of mathematics constitutes still a higher order of modeling—a third-order model!

At this philosophical conference on scientific modeling, I will make a few suggestions about the modeling of mathematics.

1 Empirical Studies of Mathematics

To study any phenomenon, a scholar or scientist must conceptualize it in one way or another. She must focus on some aspects and leave others aside. That is to say, she models it.

Mathematical knowledge and mathematical activity are observable phenomena, already present in the world, already out there, before philosophers, logicians, neuroscientists, or behavioral scientists proceed to study them.

The empirical modeling of social phenomena is a whole industry. Mathematical models, statistical models and computer models strive to squeeze some understanding out of the big data that is swamping everyone. Mathematical *activity* (in contrast to mathematical *content*) is one of these social phenomena. It is modeled by neuroscience, by logic, by history of mathematics, by psychology of mathematics, anthropology and sociology. These must use verbal modeling for phenomena that are not quantifiable—the familiar psychological and interpersonal variables of daily life, including mathematical life.

Recognizing mathematical behavior and mathematical life as empirical phenomena, we'd expect to use various different models, each focusing on a particular aspect of mathematical behavior. Some of these models might be mathematical. For such models there would be a certain reflexivity or self-reference, since the model then would be part of the phenomenon being modeled.

History, logic, neuroscience, psychology, and other sciences offer different models of mathematics, each focusing on the aspects that are accessible to its method of investigation. Different studies of mathematical life overlap, they have interconnections, but still, each works to its own special standards and criteria. Historians are historians first of all, and likewise educators, neuroscientists, and so on. Each special field studying math has its own model of mathematics.

Each of these fields has its particular definition of mathematics. Rival definitions could provoke disagreement, even conflict. Disagreement and conflict are

sometimes fruitful or instructive, but often they are unproductive and futile. I hope to convince some members of each profession that his/her viewpoint isn't the only one that is permissible. I try to do justice to all, despite the bias from a lifetime as a mathematician.

Let's look separately at four of the math-studying disciplines, and their models.

Logic. Among existing models of mathematics, the giant branch of applied logic called formalized mathematics is by far the most prestigious and successful. Being at once a model of mathematics and a branch of mathematics, it has a fascinating self-reflexivity. Its famous achievements are at the height of mathematical depth. Proudly and justifiably, it excludes the psychological, the historical, the personal, the contingent or the transitory aspects of mathematics.

Related but distinct is the recent modeling of mathematical proof in actual code that runs on an actual machine. Such programs come close to guaranteeing that a proof is complete and correct.

Logic sees mathematics as a collection of virtual inscriptions—declarative sentences that could in principle be written down. On the basis of that vision, it offers a model: formal deductions from formal axioms to formal conclusions—formalized mathematics. This vision itself is mathematical. Mathematical logic is a branch of mathematics, and whatever it's saying about mathematics, it is saying about itself—self-reference. Its best results are among the most beautiful in all of mathematics (Godel's incompleteness theorems, Robinson's nonstandard analysis).

This powerful model makes no attempt to resemble what real mathematicians really do. That project is left to others. The logician's view of mathematics can be briefly stated (perhaps over-simplified) as "a branch of applied logic".

The competition between category theory and set theory, for the position of "foundation," can be regarded as a competition within logic, for two alternative logical foundations. Ordinary working mathematicians see them as two alternative models, either of which one may choose, as seems best for any purpose.

The work of *neuroscientists* like Stanislas Dehaene (1997) is a beginning on the fascinating project of finding how and where mathematical activity takes place on the biophysical level of flesh and blood. Neuroscience models mathematics as an activity of the nervous system. It looks at electrochemical processes in the nervous system of the mathematician. There it seeks to find correlates of her mathematical process. Localization in the brain will become increasingly accurate, as new research technologies are invented. With accurate localization, it may become possible to observe activity in specific brain processes synchronized with conscious mathematical thought. Already, Jean-Pierre Changeux argues forcefully that mathematics is nothing but a brain process.

The neuroscientist's model of mathematics can be summarized (a bit over-simplified) as "a certain kind of activity of the brain, the sense organs and sensory nerves."

History of mathematics is done by mathematicians as well as historians. History models mathematics as a segment of the ongoing story of human culture Mathematicians are likely to see the past through the eyes of the present, and ask, "Was it important? natural? deep? surprising? elegant?" The historian sees

mathematics as a thread in the ever-growing web of human life, intimately interwoven with finance and technology, with war and peace. Today's mathematics is the culmination of all that has happened before now, yet to future viewpoints it will seem like a brief, outmoded stage of the past.

Many *philosophers* have proposed models of mathematics, but without explicitly situating their work in the context of modeling. Lakatos' *Proofs and Refutations* (1976) presents a classroom drama about the Descartes-Euler formula. The problem is to find the correct definition of "polyhedron," to make the Descartes-Euler formula applicable. The successive refinement by examples and counter-examples is implicitly being suggested as a model for mathematical research in general. Of course critics of Lakatos found defects in this model. His neat reconstruction overlooked or omitted inconvenient historical facts. Lakatos argued that his rational reconstruction was more instructive than history itself! This is amusing or outrageous, depending on how seriously you take these matters. It is a clear example of violating the zero'th law of modeling, which is: Never confuse or identify the model with the phenomenon!

Philip Kitcher's *The Nature of Mathematical Knowledge* (1983) sought to explain how mathematics grows, how new mathematical entities are created. He gave five distinct driving forces to account for this. Solomon Feferman (1998), in constructing the smallest system of logic that is big enough to support classical mathematics, is also offering us a model of mathematics. Emily Grosholz (2007), in focusing on what she calls "ampliative" moves in mathematical research, is modeling mathematical activity. Carlo Cellucci (2006), in arguing that plausible reasoning rather than deductive reasoning is the essential mathematical activity, is also proposing a model of mathematics. In *A Subject With No Object*, John Burgess and Gideon Rosen (1997) conclude that nominalist reconstructions of mathematics help us better understand mathematics—even though nominalism (they argue) is not very tenable as a philosophical position. This short list reflects my own reading and interests. Many others could be mentioned.

Analogous to the well-established interaction of history of science and philosophy of science, there has been some fruitful interaction between philosophy of mathematics and history of mathematics. One disappointing example was the great French number theorist Andre Weil (1978), who in his later years took an interest in history, and declared that no two fields have less in common, than philosophy of math and history of math. The philosopher-historian Imre Lakatos (1976), on the other hand, wrote that without philosophy history is lame, and without history, philosophy is blind. Or maybe it's the other way around. Each model is important, none should be ignored.

The collaboration between philosopher Mark Johnson and linguist George Lakoff is exemplary. (*Where mathematics comes from*, by Lakoff and Rafael Nunez (2000), is a major contribution to our understanding of the nature of mathematics.)

There are some eccentric, philosophically oriented *mathematicians.* We try to untangle our own and each others' actions and contributions. We don't always manage to separate the content of mathematics from the activity of mathematics, for to us they are inseparable. We aren't offering contributions to philosophy. We're not philosophers, as some philosophers politely inform us. We merely try to report

faithfully and accurately what we really do. We are kindly tolerated by our fellow-mathematicians, and are considered "gadflies" by the dominant philosophers.

William Byers (2010) introduced ambiguity as an essential aspect of mathematics, and a driving force that leads to the creation of new mathematics.

Several leading mathematicians have written accounts of their own experience in a phenomenological vein; I quote them in *How mathematicians convince each other*, one of the chapters in *Experiencing Mathematics.*

My own recent account of mathematicians' proof is another possible model of mathematics. Here it is: A mathematician possesses a mental model of the mathematical entity she works on. This internal mental model is accessible to her direct observation and manipulation. At the same time, it is socially and culturally controlled, to conform to the mathematics community's collective model of the entity in question. The mathematician observes a property of her own internal model of that mathematical entity. Then she must find a recipe, a set of instructions, that enables other competent, qualified mathematicians to observe the corresponding property of their corresponding mental model. That recipe is the proof. It establishes that property of the mathematical entity.

This is a verbal, descriptive model. Like any model, it focuses on certain specific features of the situation, and by attending to those features seeks to explain what is going on.

The discussion up to this point has left out of account the far greater part of ongoing mathematical activity—that is, schooling. Teaching and learning, education.

Teachers and educators will be included in any future comprehensive science of mathematics. They observe a lot and have a lot to say about it. Paul Ernest (1997), in his book *Social constructivism in the philosophy of mathematics*, follows Lakatos and Wittgenstein, in building his social constructivist model.

Mathematics education has urgent questions to answer. What should be the goals of math education? What methods could be more effective than the present disastrously flawed ones? Mathematics educators carry on research to answer these questions. Their efforts would be greatly facilitated by a well-established overall study of the nature of mathematics.

Why not seek for a unified, distinct scholarly activity of *mathematics studies*: the study of mathematical activity and behavior? Mathematics studies could be established and recognized, in a way comparable to the way that linguistics has established itself, as the study of mathematical behavior, by all possible methods. Institutionally, it would not interfere with or compete with mathematics departments, any more than linguistics departments impinge on or interfere with the long-established departments of English literature, French literature, Russian literature, and so on.

Rather than disdain the aspect of mathematics as an ongoing activity of actual people, philosophers could seek to deepen and unify it. How do different models fit together? How do they fail to fit together? What are their contributions and their shortcomings? What is still missing? This role for philosophy of mathematics would be higher than the one usually assigned to it.

A coherent inclusive study of the nature of mathematics would contribute to our understanding of problem-solving in general. Solving problems is how progress is made in all of science and technology. The synthesizing energy to achieve such a result would be a worthy and inspiring task for philosophy.

2 About Modeling and the Philosophy of Mathematics

Turning now to the content of mathematics rather than the activity, we are in the realm of present-day philosophy of mathematics.

Philosophers of mathematics seem to be classified by their "positions," as though philosophy of mathematics were mainly choosing a position, and then arguing against other positions. I take Stewart Shapiro's *The Oxford Handbook of Philosophy of Mathematics and Logic* (2005) as a respected representative. "I now present sketches of some main positions in the philosophy of mathematics," he writes.

Six positions appear in the table of contents, and five of them get two chapters, pro and con. Between chapters expounding logicism, intuitionism, naturalism, nominalism, and structuralism, are chapters reconsidering structuralism, nominalism, naturalism, intuitionism, and logicism. "One of these chapters is sympathetic to at least one variation on the view in question, and the other 'reconsiders'." Formalism gets only one chapter, evidently it doesn't need to be reconsidered.

"A survey of the recent literature shows that there is no consensus on the logical connections between the two realist theses or their negations. Each of the four possible positions is articulated and defended by established philosophers of mathematics."

"Taking a position" on the nature of mathematics looks very much like the vice of "essentialism"—claiming that some description of a phenomenon captures what that phenomenon "really is," and then trying to force observations of that phenomenon to fit into that claimed essence. Rival essentialisms can argue for a very long time; there is no way either can force the other to capitulate.

Such is the story of mathematical Platonism and mathematical anti-Platonism. Mark Balaguer (2001, 2013) has even published a book proving that neither of those two can *ever* be proved or disproved. "He concludes by arguing that it is not simply that we do not currently have any good arguments for or against Platonism but that we could never have such an argument." Balaguer's conclusion is correct. It is impossible in principle to *prove or disprove* any model of any phenomenon, for the phenomenon itself is prior to, independent of, our formalization, and cannot be regarded as or reduced to a term in a formal argument.

One natural model for mathematics is as story or narrative. Robert Thomas (2007, 2014) suggests such a model. Thinking of mathematical proofs or theories as stories has both obvious merits and defects. Pursuing its merits might have payoffs in research, or in teaching. That would be different from being a fictionalist—taking *the position* that mathematics IS fiction. Thomas has also suggested litigation and playing a game as models for mathematical activity.

Another natural model for mathematics is as a structure of structures (whether "ante rem" or otherwise). It is easy to see the merits of such a model, and not hard to think of some defects. Pursuing the merits might have a payoff, in benefiting research, or benefiting teaching. This would be a different matter from *being a structuralist*–taking the position that mathematics IS structure.

The model of mathematics as a formal-axiomatic structure is an immense success, settling Hilbert's first and tenth problems, and providing tools for mathematics like nonstandard analysis. It is a branch of mathematics while simultaneously being a model of mathematics, so it possesses a fascinating and bewildering reflexivity. Enjoying these benefits doesn't require one to be a formalist—to claim that mathematics IS an axiomatic structure in a formal language. Bill Thurston (2006) testifies to the needless confusion and disorientation which that formalist claim causes to beginners in mathematical research.

If a philosopher of mathematics regarded his preferred "position" as a model rather than a theory, he might coexist and interact more easily. Structuralism, intuitionism, naturalism, nominalism/fictionalism and realism/Platonism each has strengths and weaknesses as a model for mathematics. Perhaps the most natural and appealing philosophical tendency for modeling mathematics is phenomenology. The phenomenological investigations of Merleau-Ponty looked at *outer* perception, especially vision. A phenomenological approach to mathematical behavior would try to capture *an inner perception*, the mathematicians' encounter with her own mathematical entity.

If we looked at these theories as models rather than as theories, it would hardly be necessary to argue that each one falls short of capturing all the major properties of mathematics, for no model of any empirical phenomenon can claim to do that. The test for models is whether they are useful or illuminating, not whether they are complete or final.

Different models are both competitive and complementary. Their standing will depend on their benefits in practice. If philosophy of mathematics were seen as modeling rather than as taking positions, it might consider paying attention to mathematics research and mathematics teaching as testing grounds for its models.

Can we imagine these rival schools settling for the status of alternative models, each dealing with its own part of the phenomenon of interest, each aspiring to offer some insight and understanding? The structuralist, Platonist, and nominalist could accept that in the content of mathematics, even more than in heat conduction or electric currents, no single model is complete. Progress would be facilitated by encouraging each in his own contribution, noticing how different models overlap and connect, and proposing when a new model may be needed. A modeling paradigm would substitute competition for conflict. One philosophical modeler would allow the other modeler his or her model. By their fruits would they be judged.

Frege expelled psychologism and historicism from respectable philosophy of mathematics. Nevertheless, it is undeniable that mathematics is a historical entity, and that mathematical work or activity are mental work and activity. Its history and its psychology are essential features of mathematics. We cannot hope to understand mathematical activity while forbidding attention to the mathematician's mind.

As ideologies, historicism or psychologism are one-sided and incomplete, as was logicisms' reduction of mathematics to logic. We value and admire logic without succumbing to logicism. We can see the need for the history of mathematics and the psychology of mathematics, without committing historicism or psychologism.

The argument between fictionalists, Platonists and structuralists seems to suppose that some such theory could be or should be the actual truth. But mathematics is too complex, varied and elaborate to be encompassed in any model. An all-inclusive model would be like the map in the famous story by Borges—perfect and inclusive because it was identical to the territory it was mapping.

Formalists, logicists, constructivists, and so on can each try to provide understanding without discrediting each other, any more than the continuum model of fluids contradicts or interferes with the kinetic model.

3 Some Elementary Number Theory

Since nothing could be more tedious than 20 pages of theorizing about mathematics without a drop of actual mathematics, I end with an example from the student magazine *Eureka* which also appeared in the *College Mathematics Journal*. It is an amusing, instructive little sample of mathematicians' proof, and a possible test case for different models of mathematics.

A high-school exercise is to find a formula for the sum of the first n cubes. You quickly sum

$$1 + 8 + 27 + 64 + 125\ldots$$

and find the successive sums

$$1, 9, 36, 100, 225\ldots$$

You immediately notice that these are the squares of

$$1, 3, 6, 10, 15$$

which are the sums of the first n integers for

$$\mathrm{n} = 1, 2, 3, 4 \text{ and } 5.$$

If we denote the sum of the p'th powers of the integers, from the first up to the n'th, as the polynomial $S_p(n)$, which always has degree $p + 1$, then our discovery about the sum of cubes is very compact:

$$S_3(n) = [S_1(n)]^2$$

What is the reason for this surprising relationship? Is it just a coincidence?

A simple trick will explain the mystery. We will see that the sums of odd powers—the first, third, fifth, or seventh powers, and so on—are always *polynomials in the sum of the first n integers*. If you like, you could call this a "theorem."

I will give you instructions. To start, just make a table of the sums of p'th powers of the integers, with

p = 0 in the first row,
p = 1 in the second row,
p = 2 in the third row,
p = 3 in the fourth row.

Instead of starting each row at the left side of the page, start in the middle of the page, like this:

0	1	2	3	4	5
0	1	3	6	10	15
0	1	5	14	30	55
0	1	9	36	100	225

Now notice that nothing prevents you from extending these rows *to the left*—by successive *subtractions* of powers of integers, instead of adding! In the odd rows, subtracting negative values, you obtain positive entries. Here is what you get:

−5	−4	−3	−2	−1	0	0	1	2	3	4	5
15	10	6	3	1	0	0	1	3	6	10	15
−55	−30	−14	−5	−1	0	0	1	5	14	30	55
225	100	36	9	1	0	0	1	9	36	100	255

The double appearance of 0 in each row results from the fact that in the successive subtractions, a subtraction of 0 occurs between the subtractions of 1 to the p'th power and (−1) to the p'th power.

Notice the symmetry between the right and left half of each row. The symmetry of the first and third row is opposite to the symmetry of the second and fourth. These two opposite kinds of symmetry are called "odd" and "even" respectively.

(That is because the graphs of the odd and even power functions have those two opposite kinds of symmetry. The even powers 2, 4, and so on, have the same values in the negative direction as in the positive direction. For degree 2, the graph is the familiar parabola of $y = x^2$, with axis of symmetry on the y-axis. The fourth power, sixth power, and so on have more complicated graphs, but they all are symmetric with respect to the vertical axis. The graphs of the odd powers, on the other hand, (the first, third, fifth and so on), are symmetric in the opposite way, taking negative values in the negative direction (in the "third quadrant") and symmetric with respect to a point, the origin of coordinates.)

The two opposite symmetries in your little table suggest that the sum functions of the integers raised to even powers are odd polynomials, and the sums of odd powers are even polynomials.

Continuing to the left is done by *subtracting* $(-n)^p$. For the odd powers p, this is *negative*, so the result is *adding* n^p. That is the same as what you would do to continue *to the right*, adding the p'th power of the next integer. Therefore the observed symmetry for odd powers will continue for all n, and for every odd p, not just the $p = 1$ and $p = 3$ that we can read off our little table.

But surprise! The center of symmetry is not at

$$n = 0$$

but halfway between 0 and −1! Therefore, as the table shows, for odd p the polynomial $S_p(n)$ satisfies the shifted symmetry identity

$$S_p(-n) = S_p(n-1).$$

Therefore, for odd p, the squares, fourth powers and higher terms of $S_p(n)$ are even powers of $(n + 1/2)$. A sum of those *even* powers is the same thing as a sum of *all* powers of $(n + 1/2)^2$, which would be called "a polynomial in $(n + 1/2)^2$". To complete our proof, we need only show that

$$(n + 1/2)^2 = 2S_1 + 1/4.$$

Now $S_1(n)$ is very familiar, everybody knows that it is equal to

$$n(n+1)/2.$$

(There is a much-repeated anecdote about how this was discovered by the famous Gauss when he was a little boy in school.)

So then, multiplying out,

$$2S_1 = n^2 + n.$$

We do a little high-school algebra:

$$(n+1/2)^2 = n^2 + n + 1/4 = 2S_1 + 1/4,$$

so for odd p we do have S_p as a polynomial in S_1, as claimed.

I leave it to any energetic reader to work out $S_5(n)$ as a polynomial in $S_1(n)$. Since S_5 has degree 6, and S_1 is quadratic, S_5 will be cubic as a polynomial in S_1. There are only three coefficients to be calculated!

This little proof in elementary number theory *never even needed to state an axiom or hypothesis.* The rules of arithmetic and polynomial algebra didn't need to be made explicit, any more than the rules of first-order logic. Without an axiom or a hypothesis or a premise, where was the logic?

Given an interesting question, we dove right into the mathematics, and swam through it to reach the answer. We started out, you and I, each possessing our own internal model of mathematical tables, of the integers, and of polynomials in one variable. These models match, they are congruent. In particular, we agree that an odd power of a negative number is negative, and that subtracting a negative number results in adding a positive number.

I noticed that continuing the table to the left led to interesting insights. So I gave you instructions that would lead you to those insights. You followed them, and became convinced. My list of instructions is the proof!

One could elaborate this example into formalized logic. But, what for? More useful would be making it a test for competing models of mathematics (formerly "positions."). How would the structuralist account for it? The nominalist, the constructivist, the Platonist, the intuitionist? Which account is more illuminating? Which is more credible? How do they fit together? Are any of them incompatible with each other?

You may wonder, "Am I serious, asking a philosopher to take up modeling, instead of arguing for his chosen position against opposing positions?"

Yes. I am serious. The philosopher will then be more ready to collaborate with historians and cognitive scientists. The prospect for an integrated field of mathematics studies will improve.

However, such a turn is not likely to be made by many. If philosophy is all about "taking a position" and arguing against other positions, a switch from position-taking to modeling might bring a loss of standing among philosophers.

Acknowledgments I value the contributions to the understanding of mathematics made by Carlo Cellucci (2006), Emily Grosholz (2007), George Lakoff and Rafael Nunez (2000), David Ruelle (2007), Paul Livingston, Philip Kitcher (1983), Paul Ernest (1997), Mark Steiner, William Byers (2010), Mary Tiles (1991), Fernando Zalamea (2012) and Penelope Maddy. I thank Vera John-Steiner, Stephen Pollard, Carlo Cellucci and Robert Thomas (2007, 2014) for their suggestions for improving this article.

References

Balaguer, M.: Platonism and Anti-Platonism in Mathematics. Oxford University Press, Oxford (2001)

Balaguer, M.: A guide for the perplexed: what mathematicians need to know to understand philosophers of mathematics. http://sigmaa.maa.org/pom/PomSigmaa/Balaguer1-13.pdf (2013)

Burgess, J.P., Rosen, G.: A Subject With No Object. Oxford University Press, Oxford (1997)

Byers, W.: How Mathematicians Think. Princeton University Press, Princeton (2010)

Cellucci, C.: Introduction to Filosofia e matematica. In: Hersh, R. (ed.) 18 Unconventional Essays on the Nature of Mathematics, pp. 17–36. Springer, New York (2006)

Connes, A., Changeux, J.-P.: Conversations on mind, matter and mathematics, Princeton University Press, Princeton (1995)

Dehaene, S.: The Number Sense. Oxford University Press, Oxford (1997)

Ernest, P.: Social Constructivism in the Philosophy of Mathematics. SUNY Press, New York (1997)

Feferman, S.: In the Light of Logic. Oxford University Press, Oxford (1998)
Grosholz, E.: Representation and Productive Ambiguity in Mathematics and the Sciences. Oxford University Press, Oxford (2007)
Hersh, R.: On mathematical method and mathematical proof, with an example from elementary algebra. Eureka, December 2013
Hersh, R.: Why the Faulhaber polynomials are sums of even or odd powers of (n + 1/2). College Math. J. **43**(4), 322–324 (2012)
Hersh, R.: Experiencing Mathematics. American Mathematical Society, Rhode Island (2014)
Kitcher, P.: The Nature of Mathematical Knowledge. Oxford University Press, Oxford (1983)
Lakatos, I.: Proofs and Refutations. Cambridge University Press, Cambridge (1976)
Lakoff, G., Nunez, R.: Where mathematics comes from: How the embodied mind brings mathematics into being. Basic Books, New York (2000)
Ruelle, D.: The Mathematician's Brain. Princeton University Press, Princeton (2007)
Shapiro, S.: The Oxford Handbook of Philosophy of Mathematics and Logic. Oxford University Press, Oxford (2005)
Thomas, R.: The comparison of mathematics with narrative. In: van Kerkhove, B., van Bendegem, J.P. (eds.) Perspectives on Mathematical Practices, pp. 43–60. Springer, Dordrecht (2007)
Thomas, R.: The judicial analogy for mathematical publication. Paper delivered at the meeting of Canadian Society for History and Philosophy of Mathematics. May 25, Brock University, St Catherines, Ontario (2014)
Thurston, W.: On proof and progress in mathematics. In: Hersh, R. (ed.) 18 Unconventional Essays on the Nature of Mathematics, pp. 37–55. Springer, New York (2006)
Tiles, M.: Mathematics and the Image of Reason. Routledge, London (1991)
Weil, A.: History of mathematics: why and how. In: Proceedings of the International Congress of Mathematicians, Helsinki (1978)
Zalamea, F.: Synthetic Philosophy of Contemporary Mathematics. Urbanomic/Sequence Press, New York (2012)

Scientific Models Are Distributed and Never Abstract

A Naturalistic Perspective

Lorenzo Magnani

> *The biological memory records, known as engrams, differ from the external symbols, or exograms, in most of their computational properties […]. The conscious mind is thus sandwiched between two systems of representation, one stored inside the head and the other outside […]. In this case, the conscious mind receives simultaneous displays from both working memory and the external memory field. Both displays remain distinct in the nervous system.*
>
> Merlin Donald, *A Mind So Rare. The Evolution of Human Consciousness*, 2001.

Abstract In the current epistemological debate scientific models are not only considered as useful devices for explaining facts or discovering new entities, laws, and theories, but also rubricated under various new labels: from the classical ones, as abstract entities and idealizations, to the more recent, as fictions, surrogates, credible worlds, missing systems, make-believe, parables, functional, epistemic actions, revealing capacities. This article discusses these approaches showing some of their epistemological inadequacies, also taking advantage of recent results in cognitive science. I will substantiate my revision of epistemological fictionalism reframing the received idea of abstractness and ideality of models with the help of recent results related to the role of distributed cognition (common coding) and abductive cognition (manipulative).

Keywords Models · Abstract models · Idealization · Abduction · Fictions · Distributed cognition · Creativity

L. Magnani (✉)
Department of Humanities, Philosophy Section, and Computational Philosophy Laboratory, University of Pavia, Pavia, Italy
e-mail: lmagnani@unipv.it

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_13

1 Against Fictionalism

Current epistemological analysis of the role models in science is often philosophically unproblematic and misleading. Scientific models are now not only considered as useful ways for explaining facts or discovering new entities, laws, and theories, but are also rubricated under various new labels: from the classical ones, abstract entities (Giere 1988, 2009, 2007) and idealizations (Portides 2007; Weisberg 2007; Mizrahi 2011), to the more recent, fictions (Fine 2009; Woods 2010; Woods and Rosales 2010a, b; Contessa 2010; Frigg 2010a, b, c; Godfrey-Smith 2006; 2009; Suárez 2009, 2010), surrogates (Contessa 2007), credible worlds (Sugden 2000, 2009; Kuorikoski and Lehtinen 2009), missing systems (Mäki 2009; Thomson-Jones 2010), as make-believe (Frigg 2010a, b, c; Toon 2010), parables (Cartwright 2009b), as functional (Chakravartty 2010), as epistemic actions (Magnani 2004a, b), as revealing capacities (Cartwright 2009a). This proliferation of explanatory metaphors is amazing, if we consider the huge quantity of knowledge on scientific models that had already been produced both in epistemology and in cognitive science. Some of the authors mentioned above are also engaged in a controversy about the legitimacy especially of speaking of fictions in the case of scientific models.

Even if the above studies related to fictionalism have increased knowledge about some aspects of the role of models in science, I am convinced that sometimes they have also generated some philosophical confusion and it seems to me correct (following the suggestion embedded in the title of a recent paper) "to keep quiet on the ontology of models" (French 2010), and also to adopt a more skeptical theoretical attitude. I think that, for example, models can be considered fictions or surrogates, but this just coincides with a common sense view, which appears to be philosophically questionable or, at least, delusory. Models are used in a variety of ways in scientific practice, they can also work as mediators between theory and experiment (Portides 2007), as pedagogical devices, for testing hypotheses, or for explanatory functions (Bokulich 2011), but these last roles of models in science are relatively well-known and weakly disputed in the epistemological literature. In this article I will concentrate on scientific models in creative abductive cognitive processes, which Hintikka considered the central problem of current epistemological research (Hintikka 1998).

I provocatively contend that models, both in scientific reasoning and in human visual perception,[1] are neither mere fictions, simple surrogates or make-believe, nor they are unproblematic idealizations; in particular, models are never *abstract*, contrarily to the received view: of course this does not mean that the standard epistemological concept of abstract model is devoid of sense, but that it has to be

[1]In philosophical tradition visual perception was viewed very often like a kind of inference (Kant 1929; Fodor 1983; Gregory 1987; Josephson and Josephson 1994). On visual perception as model-based abduction cf. chapter five of my book (Magnani 2009); its semi-encapsulated character is illustrated in Raftopoulos (2001a, b, 2009).

considered in a Pickwickian sense. In the meantime I aim at substantiating my critique of fictionalism also outlining the first features of my own approach to the role of scientific models in terms of what I call "epistemic warfare" (see below, Sect. 4), which sees scientific enterprise as a complicated struggle for rational knowledge in which it is crucial to distinguish epistemic (for example scientific models, experiments, mathematics, etc.) from non epistemic (for example fictions, falsities, propaganda, etc.) weapons. The characteristic feature of *epistemic* weapons is that they are value-directed to the aim of promoting the attainment of scientific truth, for example through predictive and empirical accuracy, simplicity, testability, consistency, etc.

In this perspective I basically agree with the distinction between epistemic and non-epistemic values as limpidly depicted in (Steel 2010) and, substantially, with the celebration of the so-called epistemic virtues, as pragmatic—I would also add "moral"—conditions of rational truths, illustrated by Cozzo (2012). What I called "moral epistemology" (Magnani 2011, p. 274) (which for example comprehends the intrinsic "morality of sound reasoning" and is concerned with a somehow moral "commitment to the truth"), is indeed supposed to be clever in a pure way and able to foster good moral outcomes for everyone. In sum, we have to acknowledge that rationality is always intertwined with a kind of "moral" commitment: if making science as the fruit of following rules of rationality is considered central, we also have to acknowledge that a deliberate *moral* decision of following them is necessary.

I consider scientific enterprise a complicated epistemic warfare, so that we could plausibly expect to find fictions in this struggle for rational knowledge. Are not fictions typical of any struggle which characterizes the conflict of human coalitions of any kind? During the Seventies of the last century Feyerabend (1975) clearly stressed how, despite their eventual success, the scientist's claims are often far from being evenly proved, and accompanied by "propaganda [and] psychological tricks in addition to whatever intellectual reasons he has to offer" (p. 65), like in the case of Galileo. Indeed Galileo's discussions of real experiments—in the *Dialogo* but also in the *Discorsi*—become rhetorical, to confound the opponents and persuade the readers, and also to fulfil didactic needs, as contended by Naylor (1976). It is important to immediately note that another role is played by other kinds of models, for example the famous Galileo's thought experiment regarding the falling bodies, which shows the creative and constitutive—not fictional—role of cognitive models in science.[2]

The tricks that fulfil didactic or rhetorical needs are very useful and efficient, but one thing is the *epistemic* role of reasons scientist takes advantage of, such the scientific models I will illustrate in this article, which for example directly govern the path to provide a new rational intelligibility of the target systems at hand; another thing is the *extra-epistemic* role of propaganda and rhetoric, which only plays a mere—positive or negative—ancillary role in the epistemic warfare. So to

[2]I have discussed this experiment in detail in Magnani (2012).

say, these last aspects support scientific reasoning providing non-epistemic weapons able for example to persuade other scientists belonging to a rival "coalition" or to build and strengthen the coalition in question, which supports a specific research program, for example to get funds.

I am neither denying that models as idealizations and abstractions are a pervasive and permanent feature of science, nor that models, which are produced with the aim of finding the consequences of theories—often very smart and creative—are very important. I just stress that the "fundamental" role played by models in science is the one we find in the core conceptual discovery processes, and that these kinds of models cannot be indicated as fictional at all, because they are *constitutive* of new scientific frameworks and new empirical domains. In this last sense the capacity of scientific models to constitute new empirical domains and so new *knowability* is ideally related to the emphasis that epistemology, in the last century, put on the theory-ladenness of scientific facts (Hanson, Popper, Lakatos, Kuhn): in this light, the formulation of observation statements presupposes significant knowledge, and the search for new observability in science is guided by scientific modeling.[3]

Suárez (2009) provides some case studies, especially from astrophysics and concerning quantum model of measurement, emphasizing the inferential function of the supposed to be "fictional" assumptions in models: I deem this function to be ancillary in science, even if often highly innovative. Speaking of the Thomson's plum pudding model Suárez maintains that, basically "The model served an essential pragmatic purpose in generating quick and expedient inference at the theoretical level, and then in turn from the theoretical to the experimental level. It articulated a space of reasons, a background of assumptions against which the participants in the debates could sustain their arguments for and against these three hypotheses" (p. 163). In these cases the fact that various assumptions of the models are empirically false is pretty clear and so is the "improvement in the expediency of the inferences that can be drawn from the models to the observable quantities" (p. 165):[4] the problem is that in these cases models, however, are not fictions—at least in the minimal unequivocal sense of the word as it is adopted in the literary/narrative frameworks—but just the usual idealizations or abstractions, already well-known and well studied, as devices, stratagems, and strategies that lead to efficient results and that are not discarded just because they are not fake chances from the perspective of scientific rationality.[5] Two consequences derive:

[3]On this issue cf. Bertolotti (2012).

[4]It has to be added that Suárez does not conflate scientific modeling with literary fictionalizing. He clearly distinguishes scientific fictions from other kinds of fictions—the scientific ones are constrained by both the logic of inference and, in particular, the requirement to fit in with the empirical domain (Suárez 2009, 2010)—in the framework of an envisaged compatibility of "scientific fiction" with realism. This epistemological acknowledgment is not often present in other stronger followers of fictionalism.

[5]I discussed the role of chance-seeking in scientific discovery in Magnani (2007). For a broader discussion on the role of luck and chance-seeking in abductive cognition see also Bardone (2011).

– the role of models as "expediency of the inferences" in peripheral aspects of scientific research, well-known from centuries in science, does not have to be confused with the *constitutive*—in a kind of Kantian sense—role of modeling in the central creative processes, when new conceptually revolutionary perspectives are advanced [When Galileo illustrates an imaginary model—a thought experiment in this case—concerning the problem of falling bodies, he provides a kind of smart "constitutive" mental modeling (Magnani 2012)].
– models are—so to say—just models that idealize and/or abstract, but these last two aspects have to be strictly criticized in the light of recent epistemologico/cognitive literature as special kinds of epistemic actions, as I will illustrate in Sects. 2 and 3 below: abstractness and ideality cannot be solely related to empirical inadequacy and/or to theoretical incoherence (Suárez 2009, p. 168), in a static view of the scientific enterprise.

In the following sections I will concentrate my attention to the second aspect, concerning a fresh analysis of models in the light of cognitive science, which I think can help clarify some ambiguities present in the recent mainstream fictionalist epistemology of model and model-based reasoning.

Should scientific models be regarded as works of fictions? I said above that models, both in scientific reasoning and in human perception, are neither mere fictions, simple surrogates or make-believe, nor they are unproblematic idealizations; in particular, models are never abstract, contrarily to the received view. As for now we can note that, in a philosophical naturalistic framework, where all phenomena and thus also cognition, gain a fundamental eco-physical significance, models are always material objects, either when we are dealing with concrete diagrams, physical or computational models, or when we face human "mental models", which at the end "are" particular, unrepeatable, but ever-changing configurations and transformations of neural networks and chemical distributions at the level of human brains. Indeed, defending in this article an interdisciplinary approach we are simply re-engaged in one of the basic tenets of that philosophical mentality enriched by a naturalistic commitment, which acknowledges the relevance of scientific results of cognitive research.

Furthermore, if, ontologically, models are imaginary objects in the way objects of fictions are imaginary objects, I cannot see them as situated in any "location" different from the brain, so that they are imaginary in so far as they are just "mental" models. As Giere contends:

> In spite of sharing an ontology as imagined objects, scientific models and works of fiction function in different cultural worlds. One indication of this difference is that, while works of fiction are typically a product of a single author's imagination, scientific models are typically the product of a collective effort. Scientists share preliminary descriptions of their models with colleagues near and far, and this sharing often leads to smaller or larger changes in the descriptions. The descriptions, then, are from the beginning intended to be public objects. Of course, authors of fiction may share their manuscripts with family and colleagues, but this is not part of the ethos of producing fiction. An author would not be professionally criticized for delivering an otherwise unread manuscript an editor. Scientists

> who keep everything to themselves before submitting a manuscript for publication are regarded as peculiar and may be criticized for being excessively secretive (Giere 2009, p. 251).

The following sections will clarify, in a naturalistic framework, both the collective and the distributed character of scientific models.

2 Manifest Models and Perception-Action Common Coding as an Example of "On-Line" Manipulative Abduction

At the beginning of the previous section I advanced the hypothesis that models, both in scientific reasoning and in human perception, are neither mere fictions, simple surrogates or make-believe, nor they are unproblematic idealizations, and I also specifically provocatively contended that models are never *abstract* or *ideal*, contrarily to the received view: they do not live—so to say—in a kind of mysterious Popperian *World 3*. Let us deepen this second problem concerning the abstract and ideal nature of models in scientific reasoning.

First of all, within science the adopted models are certainly constructed on the basis of multiple constraints relating to the abstract laws, principles, and concepts, when clearly available at a certain moment of the development of a scientific discipline. At the same time we have to immediately stress that the same models are always *distributed* material entities, either when we are dealing with concrete diagrams or physical and computational models, or when we face human "mental models", which, as I already said in the previous section, at the end are indeed particular, unrepeatable, and ever-changing configurations and transformations of neural networks and chemical distributions at the level of human brains. In this perspective we can say that models are "abstract" only in a Pickwickian sense, that is as "mental models", shared to different extents by groups of scientists, depending on the type of research *community* at stake. This cognitive perspective can therefore help us in getting rid of some ambiguities sparked by the notion of abstractness of models.

I contend that the so-called *abstract model* can be better described in terms of what Nersessian and Chandrasekharan (2009) call *manifest model*: when the scientific collective decides whether the model is worth pursuing, and whether it would address the problems and concepts researchers are faced with, it is an internal model and it is manifest because it is shared and "[…] allows group members to perform manipulations and thus form common movement representations of the proposed concept. The manifest model also improves group dynamics" (Chandrasekharan 2009, p. 1079). Of course the internal representation presents slight differences in each individual's brain, but this does not impede that the various specific representations are clearly thought to be "abstract" insofar as they are at the same time "conceived" as referring to a unique model. This model, at a

specific time, is considered "manifest", in an atmosphere of common understanding. Nevertheless, *new* insights/modifications in the internal manifest model usually occur at the individual level, even if the approach to solve a determinate problem through the model at stake is normally shared by a specific scientific collective: the singular change can lead to the solution of the problems regarding the target system and so foster new understanding. However, new insights/modifications can also lead to discard the model at stake and to build another one, which is expected to be more fruitful and which possibly can become the new manifest model. Moreover, some shared manifest models can reach a kind of stability across the centuries and the scientific and didactic communities, like in the case of the ideal pendulum, so that they optimally reverberate the idea of high *abstractness* of scientific models.

If we comply with a conception of the mind as "extended", we can say that the mind's guesses—both instinctual and reasoned—can be classified as plausible hypotheses about "nature" because the mind grows up *together with* the representational delegations[6] to the external world that the mind itself has made throughout the history of culture by constructing the so-called cognitive niches.[7] Consequently, as I have already anticipated few lines above, not only scientific models are never abstracts/ideal, they are always distributed. Indeed, in the perspective of distributed (and embodied) cognition (Hutchins 1999) a recent experimental research (Chandrasekharan 2009, 2014; Nersessian and Chandradekharan 2009; Chandradekharan and Nersessian 2014) further provides deep and fresh epistemological insight into the old problem of abstractness and ideality of models in scientific reasoning.

The research first of all illustrates two concrete external models, as functional and behavioral approximations of neurons, one physical (in vitro networks of cultured neurons) and the other consisting in a computational counterpart, as recently built and applied in a neural engineering laboratory.[8] These models are clearly recognized as external systems—external artifacts more or less *intentionally* prepared, exactly like concrete diagrams in the case of ancient geometry—interacting with the internal corresponding models of the researchers, and they aim at generating new concepts and control structures regarding target systems. In a logico-epistemological perspective an inference which aims at generating—possibly new—hypotheses taking advantage of external models is a kind of what I have called (Magnani 2001) *manipulative abduction*. Manipulative abduction also

[6]Representational delegations are those cognitive acts that transform the natural environment in a cognitive one.

[7]The concept of cognitive niche is illustrated in detail in Odling-Smee et al. (2003). I adopted this interesting biologically oriented concept in my epistemological and cognitive research (Magnani 2009, Chap. 6), but also as an useful and synthetic theoretical tool able to clarify various puzzling problems of moral philosophy, (Magnani 2011, Chap. 4).

[8]An analysis of the differences between models in biology and physics and of the distinction between natural, concrete, and abstract models is illustrated in Rowbottom (2009); unfortunately, the author offers a description of abstract models that seems to me puzzling, and falls under some of the criticism I am illustrating in the present article.

happens when we are *thinking through doing* (and not only, in a pragmatic sense, about doing). We have to note that this kind of action-based cognition can hardly be intended as completely intentional and conscious.

As I am trying to demonstrate with the description of the above models based on common coding, I consider the interplay internal/external critical in analyzing the relation between meaningful semiotic internal resources and devices and their dynamical interactions with the externalized semiotic materiality already stored in the environment (scientific artifactual models, in this case). This external materiality plays a specific role in the interplay due to the fact that it exhibits (and operates through) its own cognitive constraints. Hence, minds are "extended" and artificial in themselves. It is at the level of that continuous interaction between on-line and off-line intelligence that I underlined the importance of manipulative abduction.

In summary, manipulative abduction, which is widespread in scientific reasoning (Magnani 2009, Chap. 1) is a process in which a hypothesis is formed and evaluated resorting to a basically extra-theoretical and extra-sentential behavior that aims at creating communicable accounts of new experiences to integrate them into previously existing systems of experimental and linguistic (theoretical) practices. Manipulative abduction represents a kind of redistribution of the epistemic and cognitive effort to manage objects and information that cannot be immediately represented or found internally. An example of manipulative abduction is exactly the case of the human use of the construction of external models in the neural engineering laboratory I have outlined above, useful to make observations and "experiments" to transform one cognitive state into another to discover new properties of the target systems. Manipulative abduction also refers to those more unplanned and unconscious action-based cognitive processes I have characterized as forms of "thinking through doing", as I have already noted above.

The external models in general offer more plasticity than the internal ones and lower memory and cognitive load for the scientist's minds. They also incorporate constraints imposed by the medium at hand that depend on the intrinsic and immanent cognitive/semiotic delegations (and the relative established conventionality) performed by the model builder(s): artificial languages, proofs, new figures, examples, computational simulations, etc.[9] It is obvious that the information (about model behavior) from models to scientists flow through perception [and not only through visualization as a mere representation—as we will see below, in the case of common coding, information also flows through "movements in the visualization [which] are also a way of generating equivalent movements in body coordinates" (Chandrasekharan 2009, p. 1076)].[10]

Perception persists in being the vehicle of model-based and motor information to the brain. We see at work that same perception that Peirce speculatively analyzed as

[9]On the cognitive delegations to external artifacts see Magnani (2009, Chap. 3, Sect. 3.6). A useful description of how specific "formats" also matter in the case of external hypothetical models and representations, and of how they provide different affordances and inferential chances, is illustrated in Vorms (2010).

[10]See also Chandradekharan (2014).

a complicated philosophical structure.[11] Peirce explains to us that some basic human model-based ways of knowing, that is *perceptions*, are abductions, and thus that they are hypothetical and withdrawable. Moreover, given the fact that judgments in perception are fallible but indubitable abductions, we are not in any psychological condition to conceive that they are false, as they are unconscious habits of inference. Hence, these fundamental—even if non scientific—model-based ways of cognizing are constitutively intertwined with inferential processes. Unconscious cognition enters these processes (and not only in the case of some aspects of perception—remind the process, in scientific modeling, of that "thinking through doing" I have just quoted above) so that in visual perception model-based cognition is typically performed in an unintentional way. The same happens in the case of emotions, which provide a quick—even if often highly unreliable—abductive appraisal/explanation of given data, which usually appears anomalous or inconsistent. It seems that, still in the light of the recent results in cognitive science I have just described, the importance of the model-based character of visual perception stressed by Peirce is intact. This suggests that we can hypothesize a continuum from construction of models that actually *emerge* at the stage of perception, where models are operating with the spontaneous application of abductive processes to the high-level model activities of more or less intentional modelers (cf. Park 2012; Bertolotti 2012), such as scientists.[12] Finally, if perception cannot be wrong, given the fact that judgments in perception are fallible but indubitable abductions, as I have just illustrated, then also these judgments should not be regarded as *fictional*.

3 Perception-Action Common Coding

The cognitive mechanism carefully exploited and illustrated in Chandrasekharan (2009, 2014) takes advantage of the notion of *common coding*,[13] recently studied in cognitive science and closely related to embodied cognition, as a way of explaining

[11]The detailed analysis of some seminal Peircean philosophical considerations concerning abduction, perception, inference, and instinct, which I consider are still important to current cognitive and epistemological research, is provided in Magnani (2009, Chap. 5).

[12]On the puzzling problem of the "modal" and "amodal" character of the human brain processing of perceptual information, and the asseveration of the importance of grounded cognition, cf. Barsalou (2008a, b).

[13]"The basic argument for common coding is an adaptive one, where organisms are considered to be fundamentally action systems. In this view, sensory and cognitive systems evolved to support action, and they are therefore dynamically coupled to action systems in ways that help organisms act quickly and appropriately. Common coding, and the resultant replication of external movements in body coordinates, provides one form of highly efficient coupling. Since both biological and nonbiological movements are equally important to the organism, and the two movements interact in unpredictable ways, it is beneficial to replicate both types of movements in body coordinates, so that efficient responses can be generated" (Chandrasekharan 2009, p. 1069): in this

the special kind of "internal-external coupling", where brain is considered a control mechanism that coordinates action and movements in the world. Common coding hypothesizes.

> […] that the execution, perception, and imagination of movements share a common representation (coding) in the brain. This coding leads to any one of these three (say perception of an external movement), automatically triggering the other two (imagination and execution of movement). One effect of this mechanism is that it allows any perceived external movement to be instantaneously replicated in body coordinates, generating a dynamic movement trace that can be used to generate an action response. The trace can also be used later for cognitive operations involving movement (action simulations). In this view, movement crosses the internal/external boundary *as movement*, and thus movement could be seen as a "lingua franca" that is shared across internal and external models, if both have movement components, as they tend to do in science and engineering (Chandrasekharan 2009, p. 1061).

Common coding refers to a representationalist account, but representation supports a motor simulation mechanism "which can be activated across different timescales—instantaneous simulation of external movement, and also extended simulations of movement. The latter could be online, that is, linked to an external movement [as in mental rotations while playing Tetris, see Kirsh and Maglio (1994)], or can be offline (as in purely imagined mental rotation)" (Chandrasekharan 2009, p. 1072). Furthermore

1. given the fact models in science and engineering often characterize phenomena in terms of bodies and particles, motor simulations are important to understand them, and the lingua franca guarantees integration between internal and external models;
2. the manipulation of the external models creates new patterns that are offered through perception to the researchers (and across the whole team, to possibly reach that shared "manifest model" I have illustrated above), and "perturbs" (through experimentation on the model that can be either intended or random) their movement-based internal models possibly leading "[…] to the generation of nonstandard, but plausible, movement patterns in internal models, which, in combination with mathematical and logical reasoning, leads to novel concepts" (cit., p. 1062);
3. this hybrid combination with mathematical and logical reasoning, and possible other available representational resources stored in the brain, offers an example of the so-called multimodality of abduction[14]: not only both data and theoretical adopted hypotheses, but also the intermediate steps between them—i.e. for example, models—can have a full range of verbal and sensory representations, involving words, sights, images, smells, etc. and also kinesthetic and motor

(Footnote 13 continued)

quoted paper the reader can find a rich reference to the recent literature on embodied cognition and common coding.

[14]On the concept of multimodal abduction see Chap. 4 of Magnani (2009).

experiences and feelings such as satisfaction, and thus all sensory modalities. Furthermore, each of these cognitive levels—for example the mathematical ones, often thought as presumptively *abstract* [does this authorize us to say they are fictional?]—actually consists in intertwined and flexible models (*external* and *internal*) that can be analogically referred to the Peircean concept of the "compound conventional sign", where for example sentential and logical or symbolic aspects coexist with model-based features. For Peirce, *iconicity hybridates logicality*: the sentential aspects of symbolic disciplines like logic or algebra coexist with model-based features—iconic. Indeed, sentential features like symbols and conventional rules[15] are intertwined with the spatial configuration, like in the case of "compound conventional signs". Model-based iconicity is always present in human reasoning, even if often hidden and implicit[16];

4. it is the perturbation I have described above that furnishes a chance for change, often innovative, in the internal model (new brain areas can be activated creating new connections, which in turn can motivate further manipulations and revisions of the external model): it is at this level that we found the scientific cognitive counterpart of what has been always called in the tradition of philosophy and history of science, scientific imagination. In a perspective that does not basically take into account the results of cognitive science but instead adopts the narrative/literary framework about models as make-believe, Toon (2010) too recognizes the role of models in perturbing mental models to favor imagination: "Without taking a stance in the debate over proper names in fiction, I think we may use Walton's analysis to provide an account of our prepared description and equation of motion. We saw […] that these are not straightforward descriptions of the bouncing spring. Nevertheless, I believe, they do represent the spring, in Walton's sense: they represent the spring by prescribing imaginings about it. When we put forward our prepared description and equation of motion, I think, those who are familiar with the process of theoretical modelling understand that they are to imagine certain things about the bouncing spring. Specifically, they are required to imagine that the bob is a point mass, that the spring exerts a linear restoring force, and so on" (p. 306).

[15]Written natural languages are intertwined with iconic aspects too. Stjernfelt (2007) provides a full analysis of the role of icons and diagrams in Peircean philosophical and semiotic approach, also taking into account the Husserlian tradition of phenomenology.

[16]It is from this perspective that—for example—[sentential] syllogism and [model-based] perception are seen as rigorously intertwined. Consequently, there is no sharp contrast between the idea of cognition as perception and the idea of cognition as something that pertains to logic. Both aspects are inferential in themselves and fruit of sign activity. Taking the Peircean philosophical path we return to observations I always made when speaking of the case of abduction: cognition is basically *multimodal*.

It is worth to note that, among the advantages offered by the external models in their role of perturbing the internal ones, there are not only the unexpected features that can be offered thanks to their intrinsic materiality, but also more neutral but fruitful devices, which can be for example exemplified thanks to the case of externalized mathematical symbols: "Apparently the brain immediately translates a positive integer into a mental representation of its quantity. By contrast, symbols that represent non-intuitive concepts remain partially semantically inaccessible to us, we do not reconstruct them, but use them as they stand" (De Cruz and De Smedt 2011). For example, it is well-known that Leibniz adopted the notation *dx* for the infinitesimals he genially introduced, and curiously called them *fictions bien fondées*, given their semantic paradoxical character: they lacked a referent in Leibnizian infinitesimal calculus, but were at the basis of plenty of new astonishing mathematical results.[17] De Cruz and De Smedt call this property of symbols "semantic opacity", which renders them underdetermined, allowing further creative processes where those same symbols can be relatively freely exploited in novel contexts for multiple cognitive aims. Semantic opacity favors a kind of reasoning that is unbiased by those intuitive aspects that possibly involve stereotypes or intended uncontrolled interpretations, typical of other less opaque external models/representations.

Peirce too was clearly aware, speaking of the model-based aspects of written proofs in deductive reasoning, that there is an "experimenting upon this image [the external model/diagram] in the imagination", where the idea that human imagination is always favored by a kind of prosthesis, the external model as an "external imagination", is pretty clear, even in case of classical geometrical deduction: "[…] namely, deduction consists in constructing an icon or diagram the relations of whose parts shall present a complete analogy with those of the parts of the object of reasoning, of experimenting upon this image in the imagination and of observing the result so as to discover unnoticed and hidden relations among the parts" (Peirce 1931–1958, 3.363).

Analogously, in the case at stake in this section, the computational model of neuronal behavior, by providing new chances in terms of control, visualizations, and costs, is exactly the peculiar tool able to favor manipulations which trigger the new idea of the "spatial activity pattern of the spikes" (Chandrasekharan 2009, p. 1067).

[17]To confront critiques and suspects about the legitimacy of the new number *dx*, Leibniz prudently conceded that *dx* can be considered a fiction, but a "well founded" one. The birth of non-standard analysis, an "alternative calculus" invented by Robinson (Robinson 1966), based on infinitesimal numbers in the spirit of Leibniz's method, revealed that infinitesimals are not at all fictions, through an extension of the real numbers system $\mathbb{R}$ to the system $\mathbb{R}^*$ containing infinitesimals smaller in the absolute value than any positive real number.

4 Epistemic Warfare: Are Scientific Models Fictions or Epistemic Weapons?

Thanks to the cognitive research I have illustrated in the previous section, we are faced with the modern awareness of what also implicitly underlies Peircean philosophical speculations I am trying to reproduce in the following paragraphs. Nature fecundates the mind because it is through a disembodiment and extension of the mind in nature (that is in such a way, so to say, "artificialized") that in turn nature affects the mind. Models are built by the mind of the scientist(s), who first delegate "meanings" to external artifacts: in other words, mind's "internal" representations are "extended" in the environment, and later on shaped by processes that are occurring through the constraints found in the external models (in the "nature" itself, Peirce would have said); that is in that external nature that consists of the "concrete" model represented by the artifact, in which the resulting aspects and modifications/movements are "picked up" and in turn re-represented and reworked in the human brain.

By the way, it is in this perspective that we can still savor, now in a naturalistic framework, the speculative Aristotelian anticipation that "nihil est in intellectu quod prius non fuerit in sensu". In such a way—that is thanks to the information that flows from the external model—the scientists' internal models are rebuilt and further refined, and the resulting modifications can easily be seen as guesses—both instinctual and reasoned, depending on the brain areas involved—that is as plausible abductive hypotheses about the external extra-somatic world (the target systems). I repeat, the process can be seen in the perspective of the theory of cognitive niches: the mind grows up together with its representational delegations to the external world that has made itself throughout the history of culture by constructing the so-called cognitive niches. In this case the complex cognitive niche of the scientific lab is an *epistemological* niche, expressly built to increase knowledge following rational methods, where "*people, systems, and environmental affordances*" (Chandrasekharan 2009, p. 1076) work together in an integrated fashion.

Even if Chandrasekharan and Nersessian's research deals with models which incorporate movement, and so does not consider models that are not based on it, it provides an useful example able to stress the distributed character of scientific models, and the true type of abstractness/ideality they possess, so refreshing these notions that come from the tradition of philosophy of science. The analysis of models as material, mathematical, and fictional—and as "abstract objects"—provided by Contessa (2010), where "a model is an actual abstract object that stands for one of the many possible concrete objects that fit the generative description of the model" (p. 228) would take advantage of being reframed in the present naturalistic perspective. The same in the case of Frigg (2010c), who contends a fictionalist view and says "Yet, it is important to notice that the model-system is not the same as its [verbal] description; in fact, we can re-describe the same system in many different ways, possibly using different languages. I refer to descriptions of this kind as model-descriptions and the relation they bear to the model-system as *p*-

representation" (pp. 257–258). Indeed, Contessa's reference to models as "actual abstract objects" and Frigg's reference to models as abstract "model-systems" would take advantage of the cognitive perspective I am presenting here: where are they located, from a naturalistic point of view? Are they mental models? If they are mental models, as I contend, this should be more clearly acknowledged.

Hence, in my perspective models cannot be considered neither abstract (in the traditional ambiguous sense) nor fictional: scientists do not have any intention to propose fictions, instead they provide models as tools that reshape a generic cognitive niche as an epistemological niche to the aim of performing a genuine struggle for representing the external world. Models, the war machines used in this struggle, which I call *epistemic warfare*, to stress the determined—strictly epistemic—dynamism of the adopted tools that are at stake, are not illusional fictions or stratagems used for example to cheat nature or swindle human beings, but just concrete, unambiguous, and well disposed tactical intermediate weapons able to strategically "attack" nature (the target systems) to further unveil its structure. Contrarily, fictions in works of fictions are for example meant to unveil human life and characters in new esthetic perspectives and/or to criticize them through a moral teaching, while fictions and stratagems in wars are meant to trick the enemy and possibly destroy the eco-human targets.

A recent study on fictions reinforces my view from a fresh epistemological perspective. Barwich (2013) analyzes the construction of models and their use in scientific reasoning by comparison with fictions, reframing the debate in terms of the presence of denoting and non-denoting elements and their "functions", so acknowledging the dynamic character of scientific cognition,[18] often overlooked by the mainstream epistemology of fictions. The examination of the role played by the so-called not-denoting elements of scientific models, which legitimated epistemological fictionalism and its fame, leads to the following conclusion: "In contrast to scientific representations, fiction is not used to serve as an explanation nor is intended to be a truthful description of the world. While scientific representations are epistemic items, proper fictions are not. In light of this, the difference between denoting and non-denoting elements is not subject to structural resemblance to a physical target system, but concerns their assigned epistemic role" (p. 367).

At this point I can conclude that scientific models are not fictions. This does not mean that fictions (and also falsities and propaganda) are not present in the scientific enterprise, as I have already anticipated in the first section of this article. To explain the features of this presence we have to introduce the reader to some issues concerning usually unnoticed moral and social aspects of the status of scientific cognition.

[18]Myself I have already emphasized the importance of taking into account the dynamic aspects of science when criticizing the epistemology of models as "missing systems": in the case of creative inferences the missing system is not, paradoxically, the one represented by the "model", but instead the target system itself, still more or less largely unknown and un-schematized, which will instead appear as "known" in a new way only after the acceptation of the research process results Magnani (2012, pp. 21–24).

I contend that epistemologists do not have to forget that various cognitive processes present a "military" nature, even if it is not evident in various aspects and uses of syntactilized human natural language and in abstract knowledge.[19] It is hard to directly see this "military intelligence"[20] in the many *epistemic* functions of natural language, for example when it is simply employed to transmit scientific results in an academic laboratory situation, or when we gather information from the Internet—expressed in linguistic terms and numbers—about the weather. However, we cannot forget that even the more abstract character of knowledge packages embedded in certain uses of language (and in hybrid languages, like in the case of mathematics, which involves considerable symbolic parts) still plays a significant role in changing the moral behavior of human collectives. For example, the production and the transmission of new scientific knowledge in human social groups not only operates on information but also implements and distributes roles, capacities, constraints and possibilities of actions. This process is intrinsically moral because in turn it generates precise distinctions, powers, duties, and chances which can create new between-groups and in-group violent (often) conflicts, or reshape older pre-existent ones.

New theoretical biomedical knowledge about pregnancy and fetuses usually has two contrasting moral/social effects, (1) a better social and medical management of childbirth and related diseases; (2) the potential extension or modification of conflicts surrounding the legitimacy of abortion. In sum, even very abstract bodies of knowledge and more innocent pieces of information enter the semio/social process which governs the identity of groups and their aggressive potential as coalitions: deductive reasoning and declarative knowledge are far from being exempt from being accompanied by argumentative, deontological, rhetorical, and dialectic aspects. For example, it is hard to distinguish, in an eco-cognitive setting, between a kind of "pure" (for example deductive) inferential function of language and an argumentative or deontological one. For example, the first one can obviously play an associated argumentative role. However, it is in the arguments traditionally recognized as fallacious, that we can more clearly grasp the military nature of human language and especially of some hypotheses reached through fallacies.

Hence, we have to be aware that science imposes itself as a paradigm of producing knowledge in a certain "decent" way, but at the same time it de facto belongs to the cross-disciplinary warfare that characterizes modernity: science more or less conflicts with other non scientific disciplines, religions, literature, magic, etc., and also implicitly orders and norms societies through technological products which impose behaviors and moral conducts. Of course scientific cognitive processes—sensu *strictu*, inside scientific groups as coalitions—also involve propaganda (and so

[19]I extendedly treated the relationship between cognition and violence in my Magnani (2011).

[20]I am deriving this expression from Thom (1988), who—in my opinion—relates "military intelligence" to the role played by language and cognition in the so-called *coalition enforcement*, that is at the level of their complementary effects in the affirmation of moralities and related conducts, and the consequent perpetration of possible violent punishments.

various kinds of "fictions" and falsities), like Feyerabend says, for instance to convince colleagues about a hypothesis or a method, but propaganda is also externally addressed to other private and public coalitions and common people, for example to get funds (a fundamental issue often disregarded in the contemporary science is the cost of producing new models) or to persuade about the value of scientific knowledge. Nevertheless the core cognitive process of science is based on avoiding fictional and rhetorical devices when the production of its own regimen of truth is at stake. Finally, science is exactly that enterprise which produces those kinds of truths which express the paradigms for *demarcating* fictions and so "irrational" or "arational" ways of knowing.

I am aware of the fact that epistemological fictionalism does not consider fictions forgery or fake, that is something "far from being execrable", instead, something "we cherish" (Frigg 2010c, p. 249), but to say that scientific and literary fictions are both "good" fictions is a bit of a theoretical oversemplification, because it is science that created, beyond literature and poetry, *new* kinds of models committed to a specific production of truth, constitutively aiming at not being fictional. I confess I cannot see how we can speak of the ideal pendulum in the same way we speak of Anna Karenina: it seems to me that we are running the risk of inadvertently opening the gates of epistemology to a kind of relativistic post-modernism *à la mode*, even if fictionalists seem to avoid this possible confusion by producing—often useful—taxonomies about the slight differences between fictions in science and in other cognitive practices.

In overall, I am convinced that introducing the word fiction in epistemology adds a modest improvement to the analysis of topics like inference, explanation, creativity, etc., but just an attractive new lexicon, which takes advantage of some seductive ideas coming for example from the theory of literary fictions. Anna Karenina and the in-vitro model[21] are very different. In actual scientific practice, a model becomes fictional only *after* the community of researchers has recognized it as such, *because* it has *failed* in fruitfully representing the target systems. In these cases a model is simply discarded. Tolstoy might have discarded the character of Anna Karenina as an inappropriate fiction for some contemporary esthetic purpose (for instance, had she failed, in her author's opinion, to veraciously represent a female member of Russia's high society at the end of XIX century), but he would have substituted her with yet another—just as fictional—character, doomed to *remain* fictional for ever.[22]

Conversely, a scientific model is recognized as fictional in a cognitive (often creative) process when it is assessed to be unfruitful, by applying a kind of *negation*

[21]Indeed, in the recent epistemological debate about fictions, even the whole "experimental systems" are reframed as "materialized fictional 'worlds'" Rouse (2009, p. 51).

[22]Giere usefully notes that "Tolstoy did not intend to represent actual people except in general terms" and that, on the contrary, a "primary function [of models in science], of course, is to represent physical processes in the real world" (Giere 2007, p. 279).

as failure (Clark 1978; Magnani 2001): it becomes fictional in the mere sense that it is falsified (even if "weakly" falsified, by failure).[23] Methodologically, negation as failure is a process of elimination that parallels what Freud describes in the case of constructions (the narratives the analyst builds about patient's past psychic life) abandoned because they do not help to proceed in the therapeutic psychoanalytic process: if the patient does not provide new "material" which extends the proposed construction, "if", as Freud declares, "[…] nothing further develops we may conclude that we have made a mistake and we shall admit as much to the patient at some suitable opportunity without sacrificing any of our authority". The "opportunity" of rejecting the proposed construction "will arise" just "[…] when some new material has come to light which allows us to make a better construction and so to correct our error. In this way the false construction drops out, as if it has never been made; and indeed, we often get an impression as though, to borrow the words of Polonius, our bait of falsehood had taken a carp of truth" (Freud 1953–1974, vol. 23, 1937, p. 262].

Similarly, for example in a scientific discovery process, the scientific model is simply eliminated and labeled as "false", because "new material has come to light" to provide a better model which in turn will lead to a new knowledge that supersedes or refines the previous one, and so the old model is buried in the necropolis of the unfruitful/dead models. Still, similarly, in the whole scientific enterprise, also a successful scientific model is sometimes simply eliminated (for example the ether model) together with the theory to which that model belonged, and so the old model is buried in yet another necropolis, that of the abandoned "historical" models, and yes, in this case, it can be plausibly relabeled as a fiction.

A conclusion in tune with my contention against the fictional character of scientific models is reached by Woods and Rosales (2010a), who offer a deep and compelling logico-philosophical analysis of the problem at stake. They contend that it is extremely puzzling to extend the theory of literary and artistic fictions to science and other areas of cognition. Whatever we say of the fictions of mathematics and science, there is "nothing true of them in virtue of which they are *literary fictions*" (p. 375). They correctly note that "Saying that scientific stipulation is subject to normative constraints is already saying something quite different from what should be said about literary stipulation":

> We also see that scientific stipulation is subject to a *sufferance* constraint, and with it to factors of timely goodness. A scientist is free to insert on his own sayso a sentence ϕ in T's model of M on the expectation that T with it in will do better than T with it not in, and subject in turn to its removal in the face of a subsequently disappointing performance by T. This is a point to make something of. Here is what we make of it:

[23]On the powerful and unifying analysis of inter-theory relationships, which involves the problem of misrepresenting models—and their substitution/adjustement—and of incompleteness of scientific representation, in terms of partial structural similarity, cf. Bueno and French (2011) and the classic (da Costa and French 2003).

- The extent to which a stipulation is held to the sufferance condition, the more it resembles a *working hypothesis*.
- The more a sentence operates as a working hypothesis, the more its introduction into a scientific theory is conditioned by *abductive considerations*.

Accordingly, despite its free standing in M, a stipulationist's ϕ in T is bound by, as we may now say, *book-end* conditions, that is to say, conditions on *admittance* into T in the first place, and conditions on its *staying* in T thereafter. The conditions on going in are broadly abductive in character. The conditions on *staying in* are broadly—sometimes very broadly—confirmational in character. Since there is nothing remotely abductive or confirmational in virtue of which a sentence is an $\mathcal{F}$-truth [fictive truth] on its author's sayso, radical pluralism must be our verdict here (Woods and Rosales 2010a, pp. 375–376).

In conclusion, after having proposed a distinction between predicates that are load-bearing in a theory and those that are not, Woods and Rosales maintain that a predicate that is not load-bearing in a theory is a *façon de parler*: "For example, everyone will agree that the predicate 'is a set' is load-bearing in the mathematical theory of sets and that 'is an abstract object', if it occurs there at all, is a *façon de parler*. 'Is an abstract object' may well be load-bearing in the philosophy of mathematics, but no work-a-day mathematician need trouble with it. It generates no new theorems for him. Similarly, 'reduces to logic' is not load-bearing in number theory, notwithstanding the conviction among logicists that it is load-bearing in mathematical epistemology" (Woods and Rosales 2010a, pp. 377–378). Unfortunately the predicate "is a fiction" is non-load-bearing, or at best a *façon de parler*, in any scientific theory. At this point the conclusion is obvious, and I agree with it, since there is no concept of scientific fiction, the question of whether it is assimilable to or in some other way unifiable with the concept of literary fiction does not arise.

Elsewhere (Magnani 2009, Chap. 3) I called the external scientific models "mimetic",[24] not in a "military" sense, as camouflaged tools to trick the hostile eco-human systems, but just as structures that mimic the target systems for epistemic aims. In this perspective I described the centrality of the so called "disembodiment of the mind" in the case of semiotic cognitive processes occurring in science. Disembodiment of the mind refers to the cognitive interplay between internal and external representations, *mimetic* and, possibly, *creative*, where the problem of the continuous interaction between on-line and off-line (for example in inner rehearsal) intelligence can properly be addressed.[25]

[24]On the related problem of resemblance (similarity, isomorphism, homomorphism, etc.) in scientific modeling some preliminary comments are provided in Magnani (2012).

[25]This distinction parallels the one illustrated by Morrison between models which idealize (mirroring the target systems) and abstract models (more creative and finalized to establish new scientific intelligibility). On this issue cf. Magnani (2012).

5 Conclusion and Future Work

In this article I have contended that scientific models are not fictions and that a naturalistic perspective can help to see abstractness and ideality of models in the concrete epistemological dimension of an eco-cognitive perspective. I have argued that also other various related epistemological approaches to model-based scientific cognition (in terms of surrogates, credible worlds, missing systems, make-believe, etc.) present severe inadequacies, which can be detected taking advantage of recent cognitive research in scientific labs and of the concept of manipulative abduction. The illustrated critique, also performed in the light of distributed cognition, offered new insight on the analysis of the two main classical attributes given to scientific models: abstractness and ideality. A further way of delineating a more satisfactory analysis of fictionalism and its discontents has been constructed by proposing the concept of "epistemic warfare", which sees scientific enterprise as a complicated struggle for rational knowledge in which it is crucial to distinguish epistemic (for example scientific models) from extra-epistemic (for example fictions, falsities, propaganda) weapons.

Other issues, I have already sketched in Magnani (2012), still need be deepened:

1. it is misleading to analyze models in science by confounding static and dynamic aspects of the scientific enterprise, or by overlooking the dynamic aspect, such as it is occurring in the case of the mainstream fictionalist epistemology: indeed the static perspective leads to an overemphasis of the possible fictional character of models because the creative/factive role of modeling is underestimated;
2. science never aimed at providing "fictions" at the basic levels of its activities,[26] so that the recent fictionalism, lacking a cognitive analysis of scientific models, does not add new and fresh knowledge about the status of models in science, and tends to obfuscate the distinctions between different areas of human cognition, such as science, religion, arts, and philosophy. In the end, "epistemic fictionalism" tends to enforce a kind "epistemic concealment", which can obliterate the actual gnoseological finalities of science, shading in a kind of debate about entities and their classification that could remind of medieval

[26]"In Sarsi [Lothario Sarsi of Siguenza is the pseudonym of the Jesuit Orazio Grassi, author of *The Astronomical and Philosophical Balance*. In *The Assayer*, Galileo weighs the astronomical views of Orazio Grassi about the nature of the comets, and finds them wanting (Galilei 1957, p. 231)]. I seem to discern the firm belief that in philosophizing one must support oneself upon the opinion of some celebrated author, as if our minds ought to remain completely sterile and barren unless wedded to the reasoning of some other person. Possibly he thinks that philosophy is a book of fiction by some writer, like the Iliad or Orlando Furioso, productions in which the least important thing is whether what is written there is true. Well, Sarsi, that is not how matters stand. Philosophy is written in this grand book, the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth" (Galilei 1957, pp. 237–238).

scholasticism. It is not certainly possible and welcome in this postmodern era of philosophy to demarcate science from non-science thanks to absolute criteria, but it is surely cognitively dangerous to systematically undermine local epistemic ways of depicting and saving the differences. The epistemological use of the so-called fictions, credible worlds, surrogate models, etc. appears theoretically suspect, but ideologically clear, if seen in the "military" framework of the academic struggle between disciplines, dominated—at least in my opinion—by a patent proliferation of "scientific" activities that just produce bare "credible" or "surrogate" models, looking aggressively for scientificity, when they actually are, at the best, kinds of fictions acting as science or, at the best, bad philosophy. I plan to devote a further study to this important topic, also intertwined with the need of reviving the old-fashioned epistemological problem of demarcation.

Acknowledgements For the instructive criticisms and precedent discussions and correspondence that helped me to develop my critique of fictionalism, I am indebted and grateful to John Woods, Shahid Rahman, Alirio Rosales, Mauricio Suárez, and to my collaborators Tommaso Bertolotti and Selene Arfini.

References

Bardone, E.: Seeking Chances. From Biased Rationality to Distributed Cognition. Springer, Heidelberg (2011)

Barsalou, L.W.: Cognitive and neural contributions to understanding the conceptual system. Curr. Dir. Psychol. Sci. **17**(2), 91–95 (2008a)

Barsalou, L.W.: Grounded cognition. Annu. Rev. Psychol. **59**, 617–645 (2008b)

Barwich, A.-S.: Science and fiction: analysing the concept of fiction in science and its limits. J. Gen. Philos. Sci. **44**, 357–373 (2013)

Bertolotti, T.: From mindless modeling to scientific models. The case of emerging models. In: Magnani, L., Li, P. (eds.) Philosophy and Cognitive Science. Western and Eastern Studies, pp. 75–104. Springer, Heidelberg/Berlin (2012)

Bokulich, A.: How scientific models can explain. Synthese **1**, 33–45 (2011)

Bueno, O., French, S.: How theories represent. Br. J. Philos. Sci. **62**, 857–894 (2011)

Cartwright, N.: If no capacities then no credible worlds. But can models reveal capacities? Erkenntnis **70**, 45–58 (2009a)

Cartwright, R.: Models: parables v. fables. Insights **1**(8), 2–10 (2009b)

Chakravartty, A.: Informational versus functional theories of scientific representation. Synthese **172**, 197–213 (2010)

Chandradekharan, S., Nersessian, N.J.: Building cognition: the construction of computational representations for scientific discovery. Cogn. Sci. 2014. doi:10.1111/cogs.12203

Chandradekharan, S.: Becoming knowledge. In: Osbek, L.M., Held, B.S. (eds.) Rational Intuition. Philosophical roots, scientific investigations, pp. 307–337. Oxford University Press, Oxford (2014)

Chandrasekharan, S.: Building to discover: a common coding model. Cogn. Sci. **33**, 1059–1086 (2009)

Clark, K.L.: Negation as failure. In: Gallaire, H., Minker, J. (eds.) Logic and Data Bases, pp. 94–114. Plenum, New York (1978)

Contessa, G.: Scientific representation, interpretation, and surrogative reasoning. Philos. Sci. **74**, 48–68 (2007)

Contessa, G.: Scientific models and fictional objects. Synthese **172**, 215–229 (2010)
Cozzo, C.: Gulliver, truth and virtue. Topoi **31**, 59–66 (2012)
da Costa, N.C., French, S.: Science and Partial Truth. A Unitary Approach to Models and Scientific Reasoning. Oxford University Press, Oxford (2003)
De Cruz, H., De Smedt, J.: Mathematical symbols as epistemic actions. Synthese **190/1**, 3–19 (2011)
Feyerabend, P.: Against Method. Verso, London (1975)
Fine, A. Fictionalism. In: Suárez M. (ed.) Fictions in Science: Philosophical Essays on Modeling and Idealization, pp. 36–19. Routledge, London (2009)
Fodor, J.: The Modularity of the Mind. The MIT Press, Cambridge (1983)
French, S.: Keeping quiet on the ontology of models. Synthese **172**, 231–249 (2010)
Freud, S.: The Standard Edition of the Complete Psychological Works of Sigmund Freud (Translated by Strachey, J. in collaboration with Freud, A. et al.). Hogarth Press, London, 1953–1974
Frigg, R.: Fiction and scientific representation. In: Frigg, R., Hunter, M.C. (eds.) Beyond Mimesis and Nominalism: Representation in Art and Science, pp. 97–138. Springer, Heidelberg (2010a)
Frigg, R.: Fiction in science. In: Woods, J. (ed.) Fictions and Models: New Essays, pp. 247–287. Philosophia Verlag, Munich (2010b)
Frigg, R.: Models and fiction. Synthese **172**, 251–268 (2010c)
Galilei, G.: The Assayer [1623]. In: Discoveries and Opinions of Galileo (Translated and edited by S. Drake), pp. 231–280. Doubleday, New York (1957)
Giere, R.N.: Explaining Science: A Cognitive Approach. University of Chicago Press, Chicago (1988)
Giere, R.: An agent-based conception of models and scientific representation. Synthese **172**, 269–281 (2007)
Giere, R.: Why scientific models should not be regarded as works of fiction. In: Suárez, M. (ed.) Fictions in Science. Philosophical Essays on Modeling and Idealization, pp. 248–258. Routledge, London (2009)
Godfrey-Smith, P.: The strategy of model-based science. Biol. Philos. **21**, 725–740 (2006)
Godfrey-Smith, P.: Models and fictions in science. Philos. Stud. **143**, 101–116 (2009)
Gregory, R.L.: Perception as hypothesis. In: Gregory, R.L. (ed.) The Oxford Companion to the Mind, pp. 608–611. Oxford University Press, New York (1987)
Hintikka, J.: What is abduction? The fundamental problem of contemporary epistemology. Trans. Charles S. Peirce Soc. **34**, 503–533 (1998)
Hutchins, E.: Cognitive artifacts. In: Wilson, R.A., Keil, F.C. (eds.) Encyclopedia of the Cognitive Sciences, pp. 126–7. The MIT Press, Cambridge (1999)
Josephson, J.R., Josephson, S.G. (eds.) Abductive Inference. Computation, Philosophy, Technology. Cambridge University Press, Cambridge (1994)
Kant, I.: Critique of Pure Reason (Translated by Kemp Smith, N. originally published 1787, reprint 1998). MacMillan, London (1929)
Kirsh, D., Maglio, P.: On distinguishing epistemic from pragmatic action. Cogn. Sci. **18**, 513–549 (1994)
Kuorikoski, J., Lehtinen, A.: Incredible worlds, credible results. Erkenntnis **70**, 119–131 (2009)
Magnani, L.: Abduction, Reason, and Science. Processes of Discovery and Explanation. Kluwer Academic, Plenum Publishers, New York (2001)
Magnani, L.: Conjectures and manipulations. Computational modeling and the extra-theoretical dimension of scientific discovery. Mind. Mach. **14**, 507–537 (2004a)
Magnani, L.: Model-based and manipulative abduction in science. Found. Sci. **9**, 219–247 (2004b)
Magnani, L.: Abduction and chance discovery in science. Int. J. Knowl.-Based Intell. Eng. **11**, 273–279 (2007)
Magnani, L.: Abductive Cognition. The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning. Springer, Heidelberg (2009)
Magnani, L.: Understanding Violence. The Interwining of Morality, Religion, and Violence: A Philosophical Stance. Springer, Heidelberg (2011)

Magnani, L.: Scientific models are not fictions. Model-based science as epistemic warfare. In: Magnani, L., Li, P. (eds.) Philosophy and Cognitive Science. Western and Eastern Studies, pp. 1–38. Springer, Heidelberg (2012)
Mäki, U.: MISSing the world. Models as isolations and credible surrogate systems. Erkenntnis **70**, 29–43 (2009)
Mizrahi, M.: Idealizations and scientific understanding. Philos. Stud. **160**(2), 237–252 (2011)
Naylor, R.: Real experiment and didactic demonstration. Isis **67**(3), 398–419 (1976)
Nersessian, N.J., Chandradekharan, S.: Hybrid analogies in conceptual innovation in science. Cogn. Syst. Res. **10**(3), 178–188 (2009)
Odling-Smee, F.J., Laland, K.N., Feldman, M.W.: Niche Construction. The Neglected Process in Evolution. Princeton University Press, Princeton (2003)
Park, W.: Abduction and estimation in animals. Found. Sci. **17**(4), 321–337 (2012)
Peirce, C.S.: Collected Papers of Charles Sanders Peirce, vols. 1–6, Hartshorne, C., Weiss, P. (eds.); vols. 7–8, Burks, A.W. (ed.). Harvard University Press, Cambridge, MA, (1931–1958)
Portides, D.P.: The relation between idealization and approximation in scientific model construction. Sci. Educ. **16**, 699–724 (2007)
Raftopoulos, A.: Is perception informationally encapsulated? The issue of theory-ladenness of perception. Cogn. Sci. **25**, 423–451 (2001)
Raftopoulos, A.: Reentrant pathways and the theory-ladenness of perception. Philos. Sci. **68**, S187–S189 (2001) (Proceedings of PSA 2000 Biennal Meeting)
Raftopoulos, A.: Cognition and Perception. How Do Psychology and Neural Science Inform Philosophy? The MIT Press, Cambridge (2009)
Robinson, A.: Non-Standard Analysis. North Holland, Amsterdam (1966)
Rouse, J.: Laboratory fictions. In: Suárez, M. (ed.) Fictions in Science: Philosophical Essays on Modeling and Idealization, pp. 37–55. Routledge, London (2009)
Rowbottom, D.P.: Models in biology and physics: What's the difference. Found. Sci. **14**, 281–294 (2009)
Steel, D.: Epistemic values and the argument from inductive risk. Philos. Sci. **77**, 14–34 (2010)
Stjernfelt, F.: Diagrammatology. An Investigation on the Borderlines of Phenomenology, Ontology, and Semiotics. Springer, Berlin (2007)
Suárez, M.: Scientific fictions as rules of inference. In: Suárez, M. (ed.) Fictions in Science: Philosophical Essays on Modeling and Idealization, pp. 158–178. Routledge, London (2009)
Suárez, M.: Fictions, inference, and realism. In: Woods, J. (ed.) Fictions and Models: New Essays, pp. 225–245. Philosophia Verlag, Munich (2010)
Sugden, R.: Credible worlds: the status of theoretical models in economics. J. Econ. Method. **7**, 1–31 (2000)
Sugden, R.: Credible worlds, capacities and mechanisms. Erkenntnis **70**, 3–27 (2009)
Thom, R.: Esquisse d'une s´emiophysique (Translated by Meyer, V.: Semio Physics: A Sketch, Addison Wesley, Redwood City, CA, 1990). InterEditions, Paris (1988)
Thomson-Jones, M.: Missing systems and the face value practice. Synthese **172**, 283–299 (2010)
Toon, A.: The ontology of theoretical modelling: Models. Synthese **172**, 301–315 (2010)
Vorms, M.: The theoretician's gambits: scientific representations, their formats and content. In: Magnani, L., Carnielli, W., Pizzi, C. (eds.) Model-Based Reasoning in Science and Technology. Abduction, Logic, and Computational Discovery, pp. 533–558. Springer, Heidelberg (2010)
Weisberg, M.: Three kinds of idealizations. J. Philos. **104**(12), 639–659 (2007)
Woods, J., Rosales, A.: Unifying the fictional. In: Woods J. (ed.) Fictions and Models: New Essays, pp. 345–388. Philosophia Verlag, Munich (2010)
Woods, J., Rosales, A.: Virtuous distortion. Abstraction and idealization in model-based science. In: Magnani, L., Carnielli, W., Pizzi, C. (eds.) Model-Based Reasoning in Science and Technology, pp. 3–30. Springer, Heidelberg (2010)
Woods, J. (ed.): Fictions and Models: New Essays. Philosophia Verlag, Munich (2010)

The Use of Models in Petroleum and Natural Gas Engineering

Kahindo Kamau and Emily Grosholz

Abstract This essay is an inquiry into the formulation of models in the science of petroleum and natural gas engineering. The underlying questions of this essay how adequate some of the fundamental models of this science really are, and what assumptions have been made as the models were created. Our claim is that a good account of the adequacy of models must be strongly pragmatist, for these questions cannot be answered properly without strict attention to human purposes. These purposes include not only the search for a better understanding of geological formations, their natural history and structure, but also classroom instruction for university students, and economically feasible extraction of petroleum and natural gas. These models include machines as well as natural formations, and so too raise the interesting question of how we (pragmatically) model machines. We claim that many of the distortions and over-simplifications in these models are in fact intentional and useful, when we examine the models in light of their pragmatic aims.

1 Introduction[1]

We start this discussion first with a series of descriptions to understand better the topic at hand. Petroleum engineering is a relatively new science which has become a central field of study in Earth and Mineral Sciences at the university level. Any introductory class in petroleum engineering will usually begin with a focus on

[1]This introduction is based on information from Dandekar (2006).

K. Kamau
Bureau of Land Management, Great Falls Field Office, 1101 15th Street,
MT 59401 Great Falls, USA
e-mail: kkamau@blm.gov

E. Grosholz (✉)
Department of Philosophy, The Pennsylvania State University, 240 Sparks Building,
University Park, PA 16802, USA
e-mail: erg2@psu.edu

E. Ippoliti et al. (eds.), *Models and Inferences in Science*,
Studies in Applied Philosophy, Epistemology and Rational Ethics 25,
DOI 10.1007/978-3-319-28163-6_14

petroleum reservoir rock and fluid properties. The term "petroleum" originated from the Greek and Latin words "petra" and "oleum", respectively, which combined form "rock oil". These two words were combined due to petroleum's location in the surface of the earth, and its common liquid state. Petroleum also refers to hydrocarbons which are compounds of carbon and hydrogen found in either liquid or vapor states. Thus the formation and location of petroleum reservoirs links them to the study of earth and mineral sciences.

The most well-received theory dealing with the formation of petroleum is known as the organic or biogenic theory. The formation process is believed to begin with subsurface generation of *kerogen*, a sedimentary organic matter generated through the decomposition of vegetable and animal organisms in sedimentary rocks. This subsurface generation is analyzed in terms of variables like pressure and temperature, and in terms of geological time scales. The rock which allows for this transformation of kerogen into petroleum is known as the *source rock*. After the formation of petroleum has occurred, hydrocarbons depart from the source rock and migrate upward through permeable beds until reaching a sealed hydrocarbon trap. This trap is an impermeable rock that allows for the accumulation of hydrocarbons and the creation of petroleum reservoirs. In an ideal reservoir, this accumulation of hydrocarbons will stay trapped by impermeable rocks until drilled for.

The typical/ideal petroleum reservoir, as seen in Fig. 1, will be found within 1600–13,000 ft below the earth's surface. At such depths it is common for temperatures and pressures to have increased tremendously. Both variables typically increase as the depth of the petroleum reservoir is increased. An ideal petroleum reservoir will be split into gas, oil, and water zones. These two phases of the hydrocarbon, and the water, separate naturally from each other due to gravity segregation.

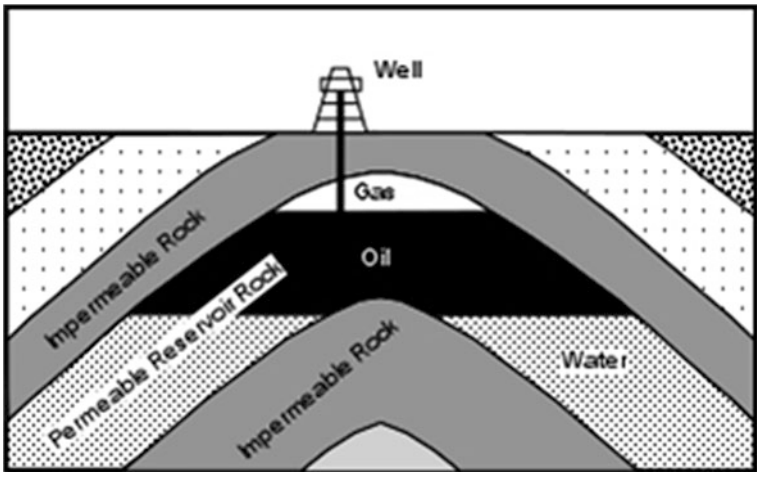


Fig. 1 Ideal petroleum reservoir ("Mineral and Energy Resources." Mineral/energy Resources. Earthsci, n.d. Web.)

2 The Progression of Models

To clarify the role that models play in the subject of petroleum engineering, we first mention the purpose of a petroleum engineer. A petroleum engineer's job is to develop ways to extract oil and natural gas from reservoirs in the earth's surface, which entails finding ways to optimize the amount of oil and gas extracted from the reservoirs efficiently and economically. Thus models must encompass a number of different dimensions, since the engineer must attend to the physical, technological and mathematics aspects of the situation. The economic aspects of the situation also strongly come into play, as well as pedagogical aspects, for this "know-how" must be conveyed to students in Earth and Mineral Sciences, even before they are sent out into the field. There are thus several types of models involved in texts devoted to petroleum engineering. The relations among these models, and the order of their introduction in the classroom, are worth studying. For example, students of petroleum engineering will begin their learning of the subject first with the physical aspects of a petroleum reservoir, before moving on to learn about mathematical models or advanced technology used in the extraction process. So as in every subject, there is a progression in the formulation of models. Typically, students begin from a certain topic or part, and progress to different topics, i.e., from physical models to mathematical models. The pattern in petroleum engineering is that physical aspects of this subject are taught prior to the mathematical. Therefore, we begin our investigation of models in petroleum engineering with physical models.

3 Representation

To investigate the use of models in this area, we must now come to a clear understanding of what models are. Models a representation of the natural world that bring scientific discourse into relation with formations or systems in nature. They rely on the use of physical science, mathematics, and empirical data to form their representations of the natural world; and they are constructed for various purposes.

In his book *Scientific Representation* (Oxford University Press 2008), Bas van Fraassen makes a three-way distinction among data models, surface models and theoretical models. Data models result from the iteration of a single operation, the interaction of an instrument with the object or system to be measured. However, such tables, plots or graphs are usually not in themselves useful for research. Highly discrete and finitary data must be reformulated into a scientifically significant form, where the data is "smoothed out" and adumbrated into a continuous, idealized, mathematically significant form. Van Fraassen calls this a surface model, and claims that it is a kind of middle term, which relates the data to a theoretical model,

which can be suitably embedded in scientific theory (Van Fraassen 2008, 166-172). This classification of models, insightful as it is, needs to be further enlarged, for models may also be simplified in terms of other human purposes, namely those of the classroom and those of the marketplace, pedagogical and economic aims.

Before we delve into the relationship that various models may have with the natural world (and the culturally constructed world of machines), we first turn to a simple example of representation to exhibit some of the issues in play. To get a better grasp of the creation of models and how they can at times mislead us, we can turn to an example given by Bas van Fraassen in his book *Scientific Representation*. Van Fraassen turns to a story related by Ernst Gombrich to show how these representations can be misleading. The story tells of two sculptors who must each create a statue of Minerva to be placed on a high pillar. The one sculptor, who is well trained in geometry and optics, creates a beautiful statue, while the other sculptor creates a disfigured and distorted statue. Once the two statues are placed on the pillar, however, everyone realized that the distorted statue looks more beautiful from a distance (Van Fraassen 2008, 12–13). This is an illustration of how models can at times be usefully misleading, for we must take into account the aim and situation of the model.

4 Simplification in Physical Models

So let us look back at Fig. 1, the Ideal Petroleum Reservoir, from an introductory online text, and ask how closely the model in Fig. 1 represents the natural phenomenon of a petroleum reservoir. A geologist and petroleum engineer analyzing this model could agree that the representation of the relationship between gas, oil, and water is only adequate as an idealization, for the figure does not include transition zones between the two phases, oil and gas, or between oil and water. The geologist might point out the structural simplification of the underground reservoir. The engineer, however, might be more concerned with the inadequacy of the representation of the above ground structures, such as the well rig. The geologist and the engineer could elaborate at length about the grounds for their severe judgments of the merits of the representation. However, the simplification in Fig. 1 can well be beneficial for students and teachers when applied at a certain period of education. To make this point, I contrast it with other models that include more detail and a truer to nature (and culture), but harder to construe.

The next two images presented, Figs. 2 and 3, are examples of a basic surface production facility and a seismic reading, respectively. The goal of a surface facility is to isolate the fluids into three components, namely oil, gas, and water. These phases must then be processed into marketable products or disposed of in an environmentally suitable manner. This picture is in its own respect simplified, but it is a much more detailed representation of the above ground occurrences on a petroleum reservoir than Fig. 1.

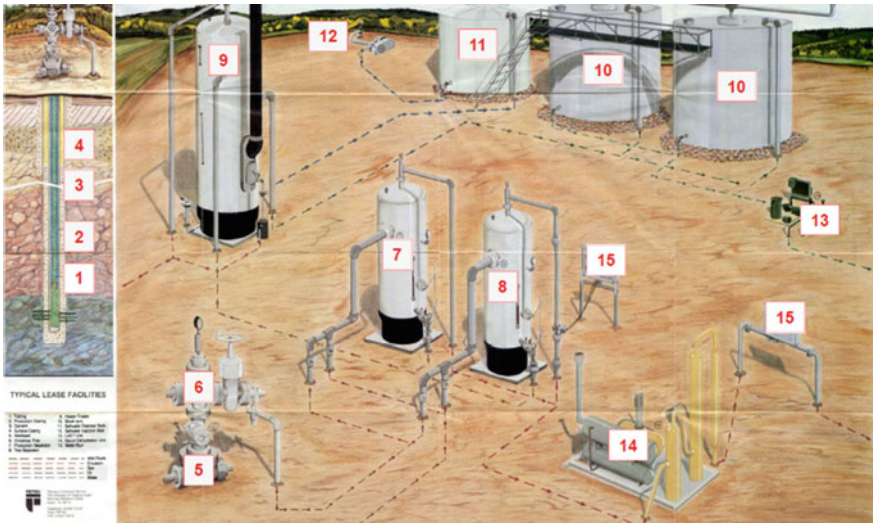


Fig. 2 Basic surface production facility (picture taken from Dr. Luis Ayala's PNG 480: Surface Production Facility notes)

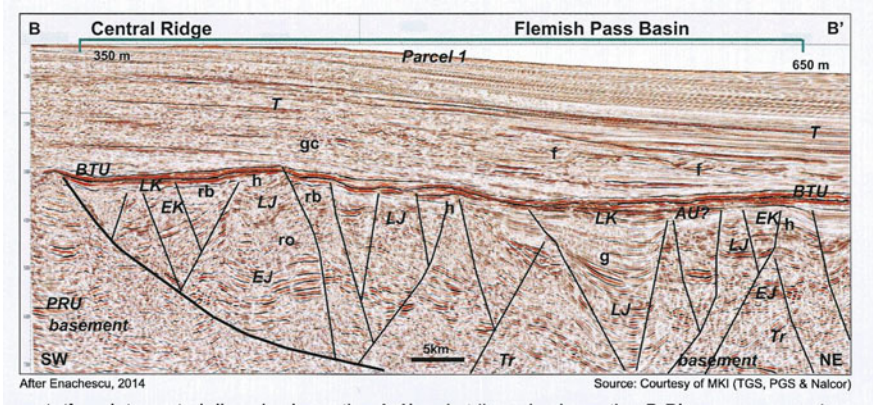


Fig. 3 Seismic reading of flemish pass basin, offshore Canada. With interpretations for faulting (picture taken from Dr. Terry Engler's GEOSC 454)

Figure 3, is an example of an actual seismic reading from below the earth's surface. In van Fraassen's terminology, it stands somewhere between a data model and a surface model. Seismic readings are accomplished in numerous ways, the most significant of which are through the use of *sources* and *receivers*. A *source* is a tool that emits a measured acoustic wave. A *receiver* is a tool which measures this acoustic wave some distance from the *source*. These acoustic waves are then reproduced as a seismic image. It takes a great deal of time to interpret seismic

images, e.g. to highlight fault lines or folds within a specific segment (as shown by the black lines in Fig. 3). A comparison of this seismic reading with Fig. 1 underlines the degree of idealization in Fig. 1.

At the stage in his or her training when a student is introduced to the first image, it provides a holistic understanding of the petroleum reservoir. That is, earlier periods of engagement with the subject, the process of understanding itself, would be obscured by the introduction of material which, at the time, might be too specific. Thus there may be good reasons why a pedagogical model simplifies a complex system. For example, it may be better in this case to summarize surface production, in a basic model of a petroleum reservoir, through the use of one rig. This rig is a denotation of all the equipment that will later be presented and explained to the student. It must be included, even in this schematic form, because it will be a vital part of a petroleum engineer's knowledge, but not at this early stage. The same goes for the idealized format of the structural formation of the petroleum geosystem.

Such an idea is highlighted in Mauricio Suárez's paper "Scientific Representation" (Suarez 2010). He argues that while philosophical conversations in the past have focused on the accuracy of scientific models, philosophers of science should now pragmatically distinguish between representation and truth. Thus he writes, "The distinction [between truth and representation] is essential to make sense of the phenomenon of scientific misrepresentation. Models often are inaccurate and misrepresent in some definite ways. This does not, however, take away any of their representational power: an imperfect and inaccurate model *M* of some target system *T* is still a model of *T*." Suárez's claim supports our current point. In the first figure we note that there is a great amount of simplification: should it be construed as a misrepresentation or an error? We can't answer that question properly without first asking what the purpose of the figure is. What is this figure trying to represent, and to what end? Figure 1, is certainly a representation of a basic petroleum reservoir; and since a petroleum reservoir is a physical phenomenon that pertains to the geological sciences, we can assume that the representation of the subsurface reservoir is the central aspect of the model.

If this assumption is acknowledged, the rig above surface could be taken as a definitive misrepresentation of a surface facility. And yet the schematic rig should still be part of the whole representation. It is still an important element of the target system, and therefore can still be used as a schematic denotation of a surface facility. The figure also simplifies its representation of a petroleum reservoir. This is seen again, in the comparison of the seismic reading in Fig. 3 and the reservoir in Fig. 1. The ideal reservoir uses a simplified interpretation of seismic images to make it easier for the student to grasp how the reservoir looks. In reality a geologist would only have seismic readings to denote the whereabouts of a reservoir, and not a simple structure map as shown in Fig. 1. It is because of the possibility of intelligent simplification and even misrepresentation in the physical models that sciences such as petroleum engineering are able to progress. Physical models use simplification as a tool to make complex phenomena into understandable material. Simplification has been used as a means for students to gain a broader understanding of their subject.

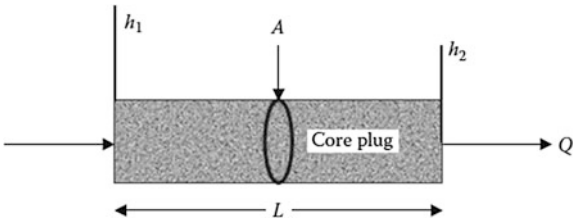


Fig. 4 Schematic of Darcy's fluid flow experiments

5 Mathematical Models[2]

Having looked at the use of iconic physical models in petroleum engineering, we can now cross over into the realm of mathematical models, specifically equational models. Our example will show how heavily mathematical models rely on an understanding of the physical world. Two of the most important rock and fluid properties that every petroleum engineer must understand are *porosity* and *permeability*. The porosity of a reservoir rock is its storage capacity—pore volume divided by total volume. The permeability of a reservoir rock is a measure of its flow capacity, or the rock's ability to transmit fluids. Solving for the permeability of a reservoir rock is done by using Darcy's equation, named after its originator.

Henry Darcy was a French civil engineer who developed the mathematical equation known today as Darcy's Law. He developed this equation while testing the flow of water through sand bags for the city of Dijon, France. Figure 4, summarizes the flow experiments that Darcy was investigating. The results of Darcy's experiments are expressed in Darcy's equation, in Eq. 1. Where Q is the volumetric flow rate through a core plug (in ft^3/s), K the proportional constant also defined as hydraulic conductivity (in ft/s), A the cross-sectional area of the core plug (in ft^2), L the length of the core plug in (ft), h_1 and h_2 represent the hydraulic head at inlet and outlet, respectively (in ft), ρ is the fluid density (in kg/ft^3), and g is the acceleration due to gravity (in ft/s^2).

$$Q = KA\frac{(h_1 - h_2)}{L} = KA\frac{dP}{dL} \tag{1}$$

$$where\ dP = \Delta h \rho g$$

This conception of flow through a core plug can then be transformed so that that it has a dependence on both permeability and fluid viscosity—creating a generalized equation that can be used for more fluids other than water. Viscosity is a measure of the fluid's resistance to flow, e.g. maple syrup has a higher viscosity

[2]This section is based heavily on Chap. 4: Absolute Permeability, in (Dandekar 2006) *Petroleum Reservoir Rock and Fluid Properties*. All figures are taken from this chapter Abhijit (2006b).

than water, and furthermore cold syrup has a higher viscosity than hot syrup. Equations 2 and 3 represent the inclusion of permeability and viscosity in Darcy's equation. In this equation K has been taken to be the ration of k/μ, where k is the permeability and μ is the viscosity of a fluid.

$$Q = -\frac{k}{\mu} A \frac{dP}{dL} \tag{2}$$

$$Q = -\frac{kA\Delta P}{\mu L} \tag{3}$$

We note that, between the articulation of Eqs. 2 and 3, an integration has taken place. However, the most important aspect for understanding Eq. 3, known as *Darcy's Law*, in this context is to understand the assumptions that had to be made for its articulation. *Darcy's Law* is an equation used extensively in petroleum engineering calculations to solve for permeability of a reservoir rock, as noted above. We will focus on one important assumption. This assumption is that the fluid flowing through the core plug has completely saturated the plug. In this core plug there is no other fluid flowing through the plug except the one being calculated for. This means that the core is 100 % saturated by this fluid. In an ideal reservoir this would be possible; however, natural reservoirs are much more complex. Even if a reservoir were 100 % saturated by one fluid, it is highly unlikely that even this fluid would have the same properties throughout the reservoir, as Darcy's law assumes. Although natural reservoirs are not as ideal as the core plug, this idealizing assumption is not very misleading, because there is still a kind of uniformity in them, in the following sense. In a natural reservoir, an engineer would find distinct layers, blocks or concentric rings that have a specific permeability. Such layers are used to an engineer's advantage. In such cases the average permeability over a series of layers is taken. An example is shown in Fig. 5.

The representation below highlights a combination that is made of three parallel layers of rock, each of which has a different permeability. From this representation of flow through a combination of parallel layers, we can see that Darcy's Law may be different for different layers in a reservoir. (There are also representations of flow

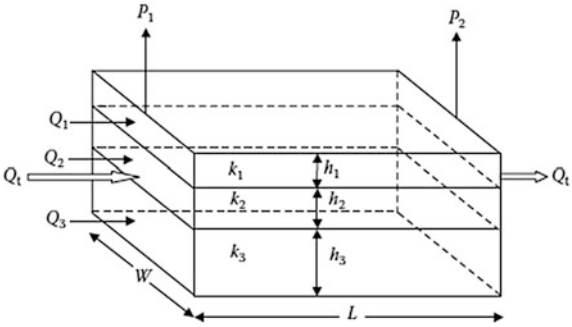


Fig. 5 A representation of flow through a parallel combination

through a series combination, though we do not discuss this further mathematical strategy in this paper. The job of a petroleum engineer is to understand the rock and fluid properties in each of the layers in these representations. Usually a summation of the flow through each of the layers represented is taken, as shown in Eq. 4. (Equations. 5 and 6 spell out the meanings of the terms.)

$$Q_t = Q_1 + Q_2 + Q_3 = \frac{k_{avg} W h_t \Delta P}{\mu L} \tag{4}$$

$$h_t = h_1 + h_2 + h_3 \tag{5}$$

$$k_{avg} = \frac{\sum_{i=1}^{n} k_i h_i}{\sum_{i=1}^{n} h_i} \tag{6}$$

These summations represent the changes in reservoir properties, such as permeability, over a whole petroleum reservoir. The summations presented in these equations offer one of the best ways to write equations for the permeability and flow rate in the layers that compose the reservoir being studied.

If petroleum reservoirs actually occurred in the nice parallel layers represented in Fig. 5, then all of the listed equations and images would seem perfectly appropriate. However, from the seismic reading shown in Fig. 3 it is clear that such parallel layerings are quite ideal. For instance, there could be parallel combinations, series combinations, and diagonal combinations. A summation of such layerings, in each specific case, could give a closer estimation of the whole reservoir's properties; but the question remains, is there really a need to build in this level of complexity? We must look at the purpose of the model, at the pragmatics of their role in human knowledge.

Before we try to answer this question, we need to review the techniques used within the petroleum industry to make permeability estimations. Usually as a well is being drilled in a certain reservoir, service companies will take core samples from different depths in the reservoir. These core samples are then carried back to laboratories and their properties are studied; during these studies equations such as Darcy's Law can be applied. Therefore companies will only know, for instance, the permeability at certain depths of a reservoir, which is not a completely precise understanding of the reservoir. Given the nature of the sampling techniques in these studies, we can conclude that companies do not need to be completely precise with their understanding of reservoirs. If necessary, the core plugs they study can be assumed to be at 100 % saturated with a fluid. From an analysis of the core permeability of each of these layers, a summation can then be taken, providing an estimation that can be used with relative confidence.

Thus even in the mathematical modeling case, we find simplification being used once again as a means to gain an understanding of a subject, in this scenario a reservoir. Simplification is therefore also necessary within mathematical modeling, as van Fraassen notes in his study of the relations between data models and surface

models. Here, however, we see that the data model itself is simplified, in addition to the "smoothing" required by the translation into equations, and that part of the motivation is practical: engineers need a model that is "good enough". We can conclude from this analysis that it is therefore not the job of the model to become an exact replica of its subject (supposing anything could ever meet that description), but rather to be a useful representation, insofar as that entails simplification and representation. A representation is therefore not completely supposed to 'be like' its target, but is supposed to accomplish whatever purpose its creator has set out for it. In the present case, the mathematical equation's purpose is to provide a "good enough" summation of a reservoir's permeability. That is, the provided estimate of permeability from Darcy's law will allow for further analysis and finally production from a reservoir that meets the expected production values for a given company. Thus economic factors bear on this situation as well.

6 Technological Models

A distinctive aspect of petroleum engineering is that it is an area of research that not only requires physical and mathematical models but also technological models. As we saw in Fig. 2, very specific equipment is used in production at a reservoir site. Equipment used at reservoir sites is modeled in terms of both physical and mathematical models. This is because the equipment is constantly interacting with the natural world that surrounds it (hence the need for iconic physical models), but is also in need of mathematical equations to help in its construction and functioning (hence the need for symbolic mathematical models). Here is one useful example: Fig. 6 is an image of a vertical separator. This separator has been cut-open for educational purposes; so once again we encounter the practical demands of pedagogy. Fluids that are produced from a reservoir are complex mixtures of hydrogen and carbon, all with different densities and other physical properties. During the production process, fluids travel from high pressures and temperatures within a reservoir to reduced pressures and temperatures at the surface facility. This period allows for the first process of phase changes to occur, such as gases evolving from liquids. These produced hydrocarbon mixtures flow from the reservoir into the wellhead—which controls production flow rate—and then into a separator.[3]

The purpose of separators is physically to separate the liquid and gas phases. The vertical separator shown above utilizes differences in gas and liquid densities to accomplish separation. It may seem to the normal human eye, and to the engineer, that this separator is quite simple, even on the inside. This is partially true. It does not take extensive knowledge of gas and liquid phase interactions to understand how gravity settling could occur due to differences in densities; and the separator is not specially complex in construction. It is mostly an empty tank that takes into

[3]Stewart and Arnold (2007).

Fig. 6 Cut-away 2 phase vertical separator (picture taken from Dr. Luis Ayala's PNG 480: Surface Production Facility notes)

account the process of gravity settling, which allows liquids to separate from gas, as well as vapor pressures, which are the pressures at which gas will evolve from liquids. Separators are therefore quite "simple" equipment. However, the process of their creation relies on both physical and mathematical modeling; and you cannot understand how they function without at least an elementary understanding of the physical and mathematical models that went into their construction. Knowledge of how gravity interacts with liquid and gas particles, and of the chemistry behind vapor pressure and its effect on fluids, is required to grasp how a vertical separator actually separates fluids. Understanding separators is therefore based on a grasp of thermodynamics, a subject concentrated on physical relationships between heat and temperature and their relation to energy and work, for thermodymics as a theory shapes the physical and mathematical models.

7 Economics

We have tried to show how models are developed in petroleum engineering research, both in the field and in the classroom. There is an iconic representation of natural phenomenon using physical modeling, an symbolic modeling of empirical data using mathematical equations, and finally equipment modeling through both

physical and mathematical modeling. These models are at times highly idealized for the benefit of students, and sometimes even use simplifications that we might call misrepresentations to achieve certain ends.

Iconic physical representation uses simplification to produce images that are easier to grasp. Mathematical equations only use a restricted amount of empirical data to produce legible, continuous solutions for various properties of reservoirs. These equations use idealized representations of the structure of reservoirs to analyze properties. Finally, technological models use physical representations and mathematical equations for the creation of equipment. From this technological modeling, equipment can then be made to interact with the natural phenomena in such a way that it will allow for maximum production from a reservoir—since this is the most common aim of a petroleum engineer.

From these considerations, we conclude modeling in petroleum engineering has important aims not only as a representation that will be most truthful to its target, but also one that will be most beneficial to its producer. For example, a petroleum engineer will not try to find the most precise reading of a reservoir's permeability, because that could take months or years; rather, he or she will aim for a moderate reading that will still allow for *economical* production. Economics therefore seems to be a leading factor in the creation of models, because it is what will indicate the most useful model to an engineer. Although we will not discuss this topic at length in this paper, we end by emphasizing its importance, because it is one of the main controlling factors in the shape of all the models here covered.

8 Conclusion

In this paper we inquire into the formulation of models in the science of petroleum and natural gas engineering. We looked at the relations among different kinds of models, and asked whether simplification might be considered misrepresentation. We concluded that, since all models are distortions, the best questions to ask are: What were the assumptions made in constructing this model? And, what were the purposes that directed its construction? The pragmatics of model construction, in order to measure reservoir properties and to create surface equipment, include the scientific interests of geologists who study oil and natural gas reservoirs, the engineers who must construct equipment to extract these products in the most efficient way, and the teachers who must train students in Earth and Mineral Science. These sometimes disparate aims explain the variety of models, and remind us that pedagogy and economics, as well as science and mathematics, play a role in shaping models.

References

Abhijit, Y.D.: Introduction. In: Introduction. Petroleum Reservoir Rock and Fluid Properties, pp. 1–4. CRC/Taylor & Francis, Boca Raton, FL (2006a) (Print)

Abhijit, Y. D.: Absolute permeability. In: Petroleum Reservoir Rock and Fluid Properties, pp. 37–48. CRC/Taylor & Francis, Boca Raton, FL (2006b) (Print)

Dandekar, A. Y.: Petroleum Reservoir Rock and Fluid Properties, Taylor & Francis, Boca Raton Florida (2006)

Stewart, M., Arnold, K.E.: Two-phase oil and gas separation. In: Surface Production Operations Design of Oil Handling Systems and Facilities, pp. 150–157. Elsevier (2007)

Suarez, M. Scientific Representation. In: Philosophy Compass, 5(1), pp. 91-101 (2010)

Van Fraassen, B.: Scientific Representation, Oxford University Press, Oxford, U.K. (2008)