Studies in Applied Philosophy, Epistemology and Rational Ethics

SAPERE

Emiliano Ippoliti *Editor*

Heuristic Reasoning

Springer

Studies in Applied Philosophy, Epistemology and Rational Ethics

Volume 16

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
Editor

Heuristic Reasoning

Springer

Editor
Emiliano Ippoliti
Department of Philosophy
Sapienza University of Rome
Rome
Italy

ISSN 2192-6255 ISSN 2192-6263 (electronic)
ISBN 978-3-319-09158-7 ISBN 978-3-319-09159-4 (eBook)
DOI 10.1007/978-3-319-09159-4

Library of Congress Control Number: 2014948769

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Reasoning at the Frontier of Knowledge: Introductory Essay

Emiliano Ippoliti

The advancement of knowledge is the big goal in human understanding. To get it, we often have to push beyond the frontier of knowledge, where our understanding dissolves and where new, strange entities appear. These require bold explorations and the consequent discoveries are not idle mind games, but crucial tools for our future life. And to have a method for carrying out these explorations is essential. Tellingly, in his famous documentary *Cosmos* and the homonymous book, Carl Sagan spent some of his most inspired words to stress this point:

> In the last few millennia we have made the most astonishing and unexpected discoveries about the Cosmos and our place within it [...]. They remind us that humans have evolved to wonder, that understanding is a joy, that knowledge is prerequisite to survival. I believe our future depends on how well we know this Cosmos in which we float like a mote of dust in the morning sky. Those explorations required skepticism and imagination both. Imagination will often carry us to worlds that never were. But without it, we go nowhere. Skepticism enables us to distinguish fancy from fact, to test our speculations [1, p. 7].

Sagan's contrasting of imagination and skepticism evokes the two main roots of logic and reasoning: ampliative reasoning, heuristics and methods for discovering on one hand, and non-ampliative reasoning, deduction, and methods for justifying and grounding our findings on the other. From these two roots have grown, branched out and borne fruit the two main traditions in logic and philosophy of science and philosophy of mathematics in particular. These traditions have seen several conflicts during the history of western scientific and philosophical thought especially on the battleground of the role of logic, reasoning and philosophy in human understanding. The latest clash was generated by the birth of mathematical logic following Frege's works. The battle is hardest fought between the orthodox view and the maverick view of philosophy of mathematics.

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

E. Ippoliti (ed.), *Heuristic Reasoning*, Studies in Applied Philosophy, Epistemology and Rational Ethics 16, DOI 10.1007/978-3-319-09159-4_1

The orthodox view is that philosophy is a meta-activity, a thinking about thinking that exists to clarify concepts, remove flaws and eradicate misunderstanding. Hence, reasoning teaches us to prevent errors, and logic is its main tool. Here logic is purely deductive, that is, a closed set of sound mechanical rules.

The maverick view claims that philosophy contributes to the hunt for new knowledge, by providing a logic and a method for its generation. Here logic is an open set of fallible rules for the generation of hypotheses from a set of data, and method is a framework for solving problems. A recent example of this view is Cellucci's revised version of the analytic method (see [2]).

Truth be told, there is a branch of the orthodox view that maintains that philosophy contributes to the advancement of knowledge, but it comes with the critical thesis that deduction and axiomatization can extend our knowledge. This is a crucial point on which the maverick view challenges the orthodoxy.

The mavericks think that deductive logic cannot genuinely extend our knowledge. They argue that no deductive rule is ampliative since the content of the conclusion is already present in its premises. According to this view, a deduction only makes explicit the information that is implicit in its premises: a deduction allows us to unfold and rewrite the information embedded in the axioms in a way that is much more understandable and testable but, *logically*, it cannot extend them. Axiomatic-deductive systems establish relations of logical dependence between *known* findings but cannot produce *new* findings.

Moreover, the relation between hypotheses and consequences, axioms and theorems, is radically different in these two views. According to a very radical maverick view, the starting point of an enquiry is not the axioms, but the consequences and the theorems. This point is expressed nicely by Hamming, who points out that this is true even in mathematics:

> The idea that theorems follow from the postulates does not correspond to simple observation. If the Pythagorean theorem were found to not follow from the postulates, we would again search for a way to alter the postulates until it was true. Euclid's postulates came from the Pythagorean theorem, not the other way. For over thirty years I have been making the remark that if you came into my office and showed me a proof that Cauchy's theorem was false I would be very interested, but I believe that in the final analysis we would alter the assumptions until the theorem was true. Thus there are many results in mathematics that are independent of the assumptions and the proof [3, pp. 86–7].

The bottom line: in the hunt for new knowledge we cannot employ axioms and deduction. Axioms are the pawns, not the queens, of our understanding, and they can be sacrificed on the chessboard of knowledge. Deductions are conservative moves: they protect your pieces and strengthen your position but do not offer ways to create lines of attack to win a match.

The orthodox view replies that axioms are the rough diamonds of our knowledge, and that deduction is the tool used to cut them. In this view the cut diamond is a new product, with new properties and relations between its parts, in other words new knowledge, so deduction is ampliative. The orthodoxy supports this claim with several arguments, such as the semi-decidability of the theories, the surprise of unexpected consequences, the need of new individuals in deduction, and the epistemic aspect of conclusions (see e.g. [4–6]). In a nutshell, these arguments set out to show that by deducing consequences we gain genuine new knowledge since the consequences (theorems) are to axioms as plants to their seeds (using a Fregean metaphor). The seeds in itself are not enough to obtain the plants, the truth of our postulates is not enough to foresee the truth of their consequences. We need an effort to obtain a deduction from given axioms—to choose and combine the premises in the appropriate way—in the same way we need a work to get a plant from its seeds. A plant is something new with respect to the seeds, so deductive consequences are new knowledge.[1] Moreover the orthodox view states that you get new plants or new properties of a plant just working on the seeds, that is on their combination and modifications. In other words, drawing deduction by relaxations, changes and combinations of axioms are ways to produce new knowledge.

The maverick view, in turn, argues that these arguments miss the big point: there is no way to logically extend our knowledge by means of deductions from axioms. The axioms are the only things needed in deductions: an axiomatic-deductive system is a closed world, unlike a plant that is an open-world that needs to interact with an environment to grow from seeds. Moreover, you don't need any kind of work or effort to get deductive consequences from a set of axioms since this task can be done mechanically by the British Museum Algorithm.

The real issue here is the definition of *new*, or *novelty*, that is what can be considered as new knowledge. On the orthodox side, establishing new logical relations between known components is regarded as new knowledge, on the maverick side only the production of an unknown component is regarded as such.

The clash between mavericks and orthodoxy, not only on the issue of new knowledge, has come to various attempts of reconciliation. For instance, recently Paolo Mancosu has tried to harmonize the two views within his 'philosophy of mathematical practice' framework (see [9]). The above Sagan's quote suggests a fruitful way to look at this problem, and also a way out to the clash. In effect, we need both ampliative and non-ampliative reasonings in the advancement of knowledge. They serve different purposes and have different roles within the same process. The ampliative reasoning offers means to produce new hypotheses capable of enlarging our understanding. The non-ampliative reasoning provides means to test and assess these hypotheses by confronting them with existing knowledge, strengthening the process of generation of hypotheses itself. In this sense, non-ampliative reasoning is useful and even necessary for the advancement of knowledge also from a maverick view point. Moreover, the work on axioms, from time to time, can produce new knowledge, since the relaxations or the changes of our postulate can be an effective heuristic move—even though it does not require to endorse a structuralist view. In particular working deductively from axioms is a means of control, a means of discovering errors in our postulates and knowledge, and learning from them—this is a big lesson from the history of set theory.

The relation between ampliative and non-ampliative reasoning can be expressed also in terms of risk-management, that is cost-benefit ratio. Basically, the non-ampliative

[1] See [7, pp. 326–332] and [8, pp. 16–72] for a detailed discussion of this point.

reasoning is a risk-aversion strategy: it aims at minimizing as much as possible the possibility of doing mistakes, but in order to reach this goal it pays a cost, that is the fact that the novel epistemic gain it offers is small or negligible. The ampliative reasoning is a risk-taking strategy: it has a potentially high cost—namely the possibility of doing bad mistakes by means of its set of fallible inferences—, which is balanced by the benefits of deep epistemic gains. This follows from the paradox of inference, which remind us that the tension between soundness and ampliativity in our reasoning cannot be dissolved.

The point is that while non-ampliative reasoning has been developed extensively in the history of philosophical and scientific thought, the same cannot be said for ampliative reasoning. One obvious reason for this is the intrinsic difficulty of producing risk-taking strategies, that is ways of reasoning at the frontier of knowledge and research. At this stage of knowledge, most of our tools for managing knowledge and solving problems vanish: the hypotheses and concepts we rely on become more and more tentative and uncertain, our knowledge-base about the objects under investigation becomes poorer and poorer, the problem-state and problem-goal can be ill-defined, the allowed 'moves' on the entities of our inquiry can be unknown or only partially known, as are the constraints on them. We really have feeble light, and most of our steps are made in darkness. Ampliative reasoning provides a way of increasing this light and so the recent resurgence of interest in it is hardly a surprise.[2]

This volume sets out to contribute to this increase and to offer ways of obtain the advancement of knowledge in this continually expanding land, populated by moving targets. But, in a sense, this difficulty is just the lesson from the 'mavericks' tradition.[3] In effect the very origin of the term 'maverick' recalls this point. It is an eponym that derives from the eccentric Texan rancher Samuel Maverick. One of his unusual traits was that he did not brand his cattle, and the noun 'maverick' was first used in 1867 to denote his unbranded cattle. Accordingly Maverick's cows turned out to be considered as outsiders, impossible to categorise by usual labels—as they were. In ampliative reasoning this feature is amplified by the fact that, quoting Bacon [22, pp. I–CXXX], "the art of discovery may improve with discoveries" ("artem inveniendi cum inventis adolescere posse"). That is, the intrinsically dynamic nature of ampliative reasoning. In effect, on one side there is the ongoing inquiry into methods for discovering, and on the other side we have that cases of discovery can be rationally evaluated, reconstructed and offered as a means of improving the 'method' of discovery itself.

The papers in the volume focus on a set of issues that are at the center of the development of ways of reasoning at the frontier of knowledge and of constructing 'methods' of discovery, such as models for revolutions and changes in paradigm, ways of treating scientific disagreement in a rational way—crucial when revolutions happen and strong disagreement can emerge inside the scientific community—, the

---

2 See in particular [2, 10–21].

3 On the role of the term and concept of "mavericks" I would like to thank David Nicholson for his valuable advice.

framework for a method of discovery and inferences for generating new knowledge, heuristics for social sciences, the use of results and findings about scientific discovery to boost funding policies capable of fostering deep impact scientific discoveries. In effect, Carlo Cellucci's and Lorenzo Magnani's papers concentrate on conceptual frameworks for scientific discovery and way of producing advancements in scientific knowledge. Emiliano Ippoliti examines four hypotheses produced in finance in order to suggest ways of generating new knowledge. Donald Gillies offers patterns for explaining the origin of revolutions and the change in paradigm in science moving from the Kuhnian approach. Dunja Seselja, Christian Strasser and Jan Willem Wieland, propose a way of treating scientific disagreement in a rational way, in order to handle disagreements that commonly emerge inside communities during revolutionary period. Tom Nickles employs results and findings about scientific discovery (e.g. the No-Free-Lunch theorems) in order to boost funding policies capable of fostering deep impact scientific discoveries or transformative research.

More specifically, the country that the mavericks are exploring lies just between the territory of the determinism of mechanical rules and the dark land of intuition. As Carlo Cellucci states in his paper *Why should the logic of discovery be revived? A reappraisal*, this country is "inhabited by heuristic procedures". And in large part they are unbranded—just like Maverick's cows. This is one of the reasons that motivates the need for a revival of the logic of discovery. Responding to the challenge why should the logic of discovery be revived? posed by Laudan in his paper *Why Was the Logic of Discovery Abandoned?* [23], Cellucci argues that the logic of discovery should be revived, on the one hand, because, as Gödel's second incompleteness theorem tells us, "mathematical logic fails to be the logic of justification, and only reviving the logic of discovery logic may continue to have an important role". On the other hand, he argues that "scientists use heuristic tools in their work, and it may be useful to study such tools systematically in order to improve current heuristic tools or to develop new ones". Following Aristotle's tenet that logic must be a tool for the method of science, Cellucci looks at inferential frameworks for scientific discovery, arguing that such frameworks are provided by a revised version of the analytic method supplemented by an open set of ampliative, non-mechanical, rules of inference: various kinds of induction and analogy, generalization, specialization, metaphor, metonymy, definition, and diagrams. Cellucci examines some of these rules in mathematical contexts and argued that they can be employed both to solve problems and to find new problems, concluding that a 'logic' of discovery is possible, without the need to call for imaginative, insightful guessing. In particular Cellucci shows how the analytic method must be distinguished from the syntetic-analytic method proposed by Aristotle. The analytic-synthetic method suffers serious limitations: above all, "it is incompatible with Gödel's incompleteness theorems". For instance there are truths of a given field "which cannot be demonstrated from those principles. Their demonstration may require principles of other fields". But, Cellucci continues, "the analytic-synthetic method requires that every truth of a given field be deducible from principles of that field. Therefore, the analytic-synthetic method is incompatible with Gödel's first incompleteness theorem".

In his paper *Are Heuristics Knowledge Enhancing? Abduction, Models, and Fictions in Science* Lorenzo Magnani focuses on 'selective' and 'creative' processes for generating hypotheses and 'cut-down' and 'fill-up' heuristics. Magnani employs an 'eco-cognitive perspective' and sets out to show that heuristics, even though non-mechanical, local and contextual, is the only means to extend our knowledge, defending the idea that its outcomes are not fictional. More specifically, Magnani focusses on the abduction as a means to produce new knowledge and he critically evaluates the status of abductive inferences by defining it as "very controversial". In effect, the examination of abduction requires answering to a series of questions: does "abduction involve only the generation of hypotheses or also their evaluation", the "criteria for the best explanation in abductive reasoning are epistemic, pragmatic, or both", or again does "abduction preserve ignorance or extend truth or both". Magnani provides an answer based on the so-called ignorance-preservation characterization of abduction, "contrasted with its knowledge enhancing capacity, such as it is expressed by its heuristic features" and he maintains that "even if, certainly, abductive reasoning can be considered a response to an ignorance-problem, nevertheless, through abduction, knowledge can be enhanced".

In his paper *Heuristic Appraisal at the Frontier of Research* Thomas Nickles shows how better abd better understanding of scientific discoveries can improve the funding and support of research. In particular, he deals with the problem of heuristic appraisal (HA) at the frontier of research and its impact in policy. The heuristic appraisal is the "identification and evaluation of hints and clues that can provide direction to inquiry in the sometimes large gap between the extremes of complete knowledge and complete ignorance". Nickles contrasts heuristic appraisal with the traditional confirmational appraisal (CA): HA is prospective, "directed toward possible future developments, future opportunities", while CA is "retrospective, based on past performance". Moving from Meno's aporia and the No Free Lunch Theorems [24, 25], ha argues that only a local, domain-specific view on a 'logic of discovery' is possible. In particular, he maintained that once problem constraints and HA hints are exhausted, we can only proceed blindly, by trial and error: in this sense he states that all genuinely new knowledge is produced by an undirected variation-and-selection process. Then he applied HA to the decision-making in the funding of pioneering research and suggested ways to stimulate 'transformative research' policies—that is "changes that challenge current understandings, either by undermining them or by opening up new areas of investigation that current views give us no reason to anticipate and that may even have been inconceivable before". Hence these change are not breakthroughs "in the sense of applications of already extant science and technology". In particular Nickles is interested in understanding how it is possible to "speed up both basic and translational scientific research without major new financial investment". This requires solving what he labels the *policy problem*, that is the fact that most funding agencies (especially in government) are designed "to discourage transformative HA recognition or to undervalue it in the interests of short-term accounting". Nickles argues that this collides with the fact that "history informs us that the innovation timescale is typically an order of magnitude or more larger than the de facto accounting timescale imposed by such requirements as 'broader impacts'. There is too much risk-avoidance, too much emphasis on quasi-guaranteed results". Thus, Nickels argues for an increased weight to heuristic appraisal and less weight to confirmational appraisal. He examines several models for fostering research activity in general, and some for encouraging transformative research: the prizes/awards model, the Linus Pauling model, the NSF model, the DARPA model, the 'triple helix' model, the Rockefeller Foundation model. In the end, his contention is that it is not possible to "realistically plan (or fund) a successful revolution, and it is difficult to identify something as a revolution even while it is occurring, at least until it has been largely accomplished. Typically, what is accomplished is not what the instigators may originally have expected. The more profound the revolution, the more difficult it is to appreciate the likely outcome and its far-reaching implications in advance". Hence, he offers a 'general policy advice for the longer term', which "focus on removing barriers and creating general opportunities rather than on pretending to give specific directions to the specialists in their domains". In the end, Nickels endorses a scenario-planning approach to funding transformative research, that is a 'as-if thinking' that involves challenging established truth, and which requires to retain an open future (in contrast with the end-of-history view).

The dynamic of scientific revolution is the center of Donald Gillies' paper *Why do Scientific Revolutions begin?*, which starts from a critique of the Kuhnian 'Build-up of Anomalies' model and presents two patterns for scientific revolutions: the tech-fist and the tech-last model. In the 'tech first' model, advances in technology come first, enabling new observations and experiments, which result in discoveries that give rise to the scientific revolution. In order to better illustrate this model Gillies provides a negative example, that is an example of what was not actually the beginning of a scientific revolution: Galileo's telescopic discoveries. Gillies notes that "the discoveries, which Galileo made in such a short space of time with his new instrument, were truly remarkable". In this example technological developments "lead to new instruments, and, with the help of these, a number of striking new discoveries are made". Gillies states that the 'tech first' pattern is, in some cases, what stimulates the beginning of a scientific revolution, and explicitly replaces a build up of anomalies theory with a build up of new discoveries theory. He offers an extensive discussion of the beginning of the chemical revolution as an example of the first model, and shows that the build up of discoveries concerning new gases and their properties gave rise to the chemical revolution. In the 'tech last' model, urgent practical hard-to-solve problems, stimulate solutions by changing the paradigm and advances in tech occur as a consequence of the scientific revolution. He illustrates these features of the model by an example drawn from the history of medicine, that is the Germ Theory of Disease—one of the big revolutions in medicine started about 1865 and largely succeeded by about 1885. This revolution ended up establishing the germ theory of disease as a new paradigm for medicine, and brought antisepsis into the practice of surgery. Tellingly, Gillies argues that the distinction between tech first and tech last is important, but many scientific revolutions can stems from an interplay of both patterns, for “partly because scientific revolutions very often have different phases, and partly because it is often difficult to decide how exactly a scientific revolution should be characterised”.

In the paper *Withstanding Tensions: Scientific Disagreement and Epistemic Tolerance* Dunja Seselja, Christian Strasser, and Jan Willem Wieland deal with the issue of disagreement in science and how this can be shown to be rational, looking at similarities to epistemic paradoxes. They offer the solution of epistemic tolerance: a normative framework allowing scientists to continue to pose a fruitful challenge, without dismissing their opponents’ stance as epistemically futile. More specifically Seselja e Strasser move from a definition of rational scientific disagreement as disagreement on some issue plus “reasons to suppose that the stance of each participant is the result of a rational deliberation”. Then, they distinguish between the internal recognition of disagreements—by the participants in a debate, and the external one—and the outside observer (e.g. a philosopher or a historian of science): each of these kinds generates certain tensional situations. They argue that scientific controversies often involve such rational disagreements, and set out to show how scientists can tentatively recognize that their disagreement is rational: namely, on the basis of content- and form-based indices. This leads them to consider the normative question about what kind of “epistemic stance a scientist should have who has recognized she may be involved in a rational disagreement”. They show that the tension characterizing rational disagreements has properties similar to epistemic paradoxes and to the notion of toleration—as it is used in ethics and politics. Hence, they introduce the notion of epistemic toleration to answer this normative question, by providing a normative framework that allows scientists to keep on posing a fruitful challenge and at the same time taking their opponents’ stance as epistemically reasonable.

In her paper *Heuristics as Methods: Validity, Reliability and Velocity* Anna Grandori deals with the application of heuristics to economic problems, showing the importance and performance implications of rational heuristics in economics, in particular decisions in which resources are scarce and performance important. She argues that there are areas where those heuristics can be applied “very fast, and errors reduced drastically”. Grandori reviews research on innovative economic and organizational decision-making processes using epistemological criteria, and shows that an array, or better a portfolio of effective and ‘rational’ heuristics can be specified—different from the repertory of ‘behavioral’, potentially ‘biasing’, heuristics usually considered. Two case studies of innovative decision making under uncertainty are examined in the paper: a new product development (a major project for reducing traffic pollution) and entrepreneurial decision making (protocol analyses of financial angels’ investing decisions). Grandori sets out to show that the heuristics applied do resemble more the ‘slow and safe’ heuristics of scientific discovery, rather than the ‘fast and frugal’ heuristics of everyday life. Then, she discusses a third case study of decision making on military flights, addressing the question of whether heuristics can be ‘fast and rational’ simultaneously. She argues that results suggest that they can, and “help in identifying the rather unexplored rational heuristics sustaining ‘highly reliable’ action under risk”.

Economics, and finance in particular, is the starting point of Emiliano Ippoliti's paper *Dynamic generation of hypotheses: Mandelbrot, Soros and Far-From-Equilibrium.* In order to investigate ways of generating hypotheses Ippoliti examines four hypotheses for dealing with the behavior of stock market prices, arguing that the generation of new hypotheses draws on a preliminary bottom-up, verbal, non-formal conceptualization, and maintained that this is the only way to incorporate the domain-specific features of the subject. In particular he examine the construction process of one hypothesis for stock market prices behavior, that is the far-from-equilibrium hypothesis. In order to do this he analyzes the generation of the hypotheses that preceded the far-from-equilibrium hypothesis. First of all, he considers the Efficient Market Hypothesis (EMH, see [26, 27]), pointing at its main vulnerability, the idea of 'equilibrium', which does not enable us to explain booms-and-busts, or at least their frequency. Then, he examines the Fractal Market Hypothesis, which offers a new interpretation of the data and shows new properties of financial markets, undermining the effectiveness of the notion of equilibrium. He argues that even though it does not explain the reasons for these properties and does not offer predictions that can be put to use–due to the sensitivity to initial conditions–it generates new mathematics and explain to us when we can expect markets to be stable. Hence, he analyses the Reflexive Market Hypothesis (see [28]), which has received little scholarly attention but offers a cogent, qualitative explanation of several properties identified but not explained by the Fractal Market Hypothesis. This hypothesis draws on the distinction between endogenous and exogenous forces in the behavior of prices and it enables us to explain boom-and-bust and crashes. In the end, he approach the Far-form-Equilibrium Hypothesis, showing how it relies on the distinction between exogenous and endogenous forces and does develop a means to forecast crashes and bubbles, for instance the so called flash-crashes (e.g. [28, 29]). The main point of this paper is to show how the means of generating these hypotheses is essential to assessing their efficiency and plausibility. More specifically he argues that in formulating a hypothesis, a selection of features of SMP is made for incorporation in a theory. This selection may be expressed mathematically in most of the cases. An examination of these means of generation can show us why some of these hypotheses are successful and efficient and some not, and can also shed light on the extent to which a particular hypothesis can be usefully applied. Thus Ippoliti argues that the study of the means of generation of hypotheses offers us a guide to formulating new hypotheses in a reliable and cogent fashion. More specifically he states that the generation of a new hypothesis has to draw on a preliminary verbal conceptualization (a discourse) on a specific subject, that is a verbal and non-formal description of it, which establishes the entities to investigate, their properties and relations, and a set of variables that affect them. This is a bottom-up process and it is the only way to incorporate the (domain) specific features of the subject in a plausible representation of it, which can possibly end up in a mathematical theory. Thus his thesis is that generation of new hypotheses and, possibly, new mathematics stems from a preliminary verbal reasoning and conceptualization, which delimitate the variables and the features of a phenomenon.

References

1. Sagan, C.: Cosmos. Random House, New York (1980)
2. Cellucci, C.: Rethinking Logic. Logic in Relation to Mathematics, Evolution, and Method. Springer, New York (2013)
3. Hamming, R. 1980. The Unreasonable Effectiveness of Mathematics. Am. Math. Monthly **87**(2), 81–90 (1980)
4. Dummett, M.: Frege. Philosophy of Mathematics. Duckworth, London (1991)
5. Hintikka, J.: Logic, Language-Games and Information. Oxford University Press, Oxford (1973)
6. Rota, G.C.: Indiscrete Thoughts. Birkhäuser, Boston (1997)
7. Cellucci, C.: Le ragioni della logica. Laterza, Roma-Bari (1998)
8. Cellucci, C.: Filosofia e Matematica. Laterza, Roma-Bari (2002)
9. Mancosu, P.: The Philosophy of Mathematical Practice. Cambridge University Press, Cambridge (2008)
10. Lakatos, I.: Proofs and Refutations. Cambridge University Press, Cambridge (1976)
11. Nickles, T. (ed.): Scientific Discovery, Logic and Rationality. Reidel, Dordrecht (1980)
12. Nickles, T. (ed.): Scientific Discovery: Case Studies. Reidel, Dordrecht (1980)
13. Nickles, T., Meheus, J. (eds.): Models of Discovery and Creativity. Springer, New York (2009)
14. Cellucci, C., Gillies, D.: Mathematical Reasoning and Heuristics. College Publications, London (2005)
15. Gillies, D. (ed.): Revolutions in Mathematics, pp. 353, xi. Oxford University Press, Oxford (1992)
16. Kantorovich, A.: Scientific Discovery: Logic and Tinkering. Suny Press, New York (1993)
17. Groshoz, E.: Representation and Productive Ambiguity in Mathematics and the Sciences. Oxford University Press, New York (2007)
18. Grosholz, E., Breger, H.: The Growth of Mathematical Knowledge. Springer, New York (2000)
19. Magnani, L.: Abduction, Reason, and Science. Processes of Discovery and Explanation. Kluwer Academic, New York (2001)
20. Magnani, L., Nersessian, N.J. (eds.): Model-Based Reasoning. Scientific Discovery, Technological Innovation, Values. Kluwer Academic, New York (2002)
21. Magnani, L., Nersessian, N.J., Thagard, P. (eds.): Model-Based Reasoning in Scientific Discovery. Kluwer Academic, New York (1999)
22. Bacon, F.: (1620) In: Jardine, L., Silverthorne, M. (eds.) Novum Organon. Cambridge University Press, Cambridge (2000)
23. Laudan, L.: Why Was the Logic of Discovery Abandoned? In: Nickles, T. (ed.) Scientific Discovery, Logic, and Rationality, pp. 173–183. Reidel, Dordrecht (1980)
24. Wolpert, D.H., Macready, W.G.: No free lunch theorems for search. Technical Report SFI-TR-95-02-010 (Santa Fe Institute) (1995)
25. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. IEEE Trans. Evol. Comput. **1**, 67 (1997)
26. Friedman, M.: The methodology of positive economics. In: Caldwell, B. (ed.) Appraisal and Criticism in Economics: A Book of Readings. Allen & Unwin, London (1953) (reprinted in 1984)
27. Fama, E.F.: Efficient capital markets: a review of theory and empirical work. J. Finance **25**(2), 383–417 (1970)
28. Soros, G.: The Alchemy of Finance. Wiley, New York (1987)
29. Sornette, D.: Why Stock Markets Crash: Critical Events in Complex Financial Systems. Princeton University Press, Princeton (2003)
30. Sornette, D., Filimonov, V.: Quantifying reflexivity in financial markets: towards a prediction of flash crashes. Phys. Rev. E **85**(5), 056108 (2012)

Why Should the Logic of Discovery Be Revived? A Reappraisal

Carlo Cellucci

Abstract Three decades ago Laudan posed the challenge: Why should the logic of discovery be revived? This paper tries to answer this question arguing that the logic of discovery should be revived, on the one hand, because, by Gödel's second incompleteness theorem, mathematical logic fails to be the logic of justification, and only reviving the logic of discovery logic may continue to have an important role. On the other hand, scientists use heuristic tools in their work, and it may be useful to study such tools systematically in order to improve current heuristic tools or to develop new ones. As a step towards reviving the logic of discovery, the paper follows Aristotle in asserting that logic must be a tool for the method of science, and outlines an approach to the logic of discovery based on the analytic method and on ampliative inference rules.

1 Introduction

In the last century, scientific discovery has been generally held to be beyond the scope of rationality. The received view has been that scientific discovery is the unique product of intuition.

Thus Planck states that the creative scientist "must have a vivid intuitive imagination, for new ideas are not generated by deduction, but by an artistically creative imagination" [1, p. 109]. Einstein states that "there is no logical path to" the basic laws of physics, "only intuition, resting on sympathetic understanding of experience, can reach them" [2, p. 226]. Indeed, only "the intuitive grasp of the essentials of a large complex of facts leads the scientist" to a "basic law, or several such basic laws" [3, p. 108]. Reichenbach states that "the act of discovery escapes logical analysis" [4, p. 231]. The "scientist who discovers a theory" cannot "name a method by means of which he found the theory," and can only say "that he saw

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

E. Ippoliti (ed.), *Heuristic Reasoning*, Studies in Applied Philosophy, Epistemology and Rational Ethics 16, DOI 10.1007/978-3-319-09159-4_2

intuitively which assumption would fit the facts" (ibid., p. 230). Discovery "is a process of intuitive guessing and cannot be portrayed by a rational procedure controlled by logical rules" since "there are no such rules" [5, p. 434].

The received view has its foundation in the romantic theory of the scientific genius going back to Novalis, who states that discoveries "are leaps—(intuitions, resolutions)" and products "of the genius—of the leaper *par excellence*" [6, p. 28]. The genius brings forth numerous living thoughts, and "whoever is able to bring forth numerous living thoughts, is called a genius" (ibid., p. 194). (See also [7]).

Contrary to the romantic theory of the scientific genius, it can be argued that discovery can be pursued through rational procedures, specifically heuristic procedures. Although the latter offer no complete guarantee to reach a solution, they restrict the search space thus easing the search for a solution.

The importance of heuristic procedures is stressed by Lakatos who, in his early work, seems genuinely interested in "the logic of mathematical discovery" [8, p. 4]. But the heuristic rules he provides are not genuine discovery rules (see [9]). On the other hand, in his later work, by 'logic of discovery' or 'methodology' Lakatos "no longer means rules for arriving at solutions, but merely directions for the appraisal of solutions already there. Thus" logic of discovery or "methodology is separated from heuristics" [10, I, p. 103, Footnote 1].

The aim of this paper is to reexamine the question of discovery, seen as pursued through rational procedures. Specifically, the paper considers some of the ways in which the question of discovery has been dealt with in the past. It follows Aristotle in asserting that logic must be a tool for the method of science, and outlines an approach to the logic of discovery based on the analytic method and on ampliative inference rules, that is, inference rules where the conclusion is not contained in the premises.

2 Attempts to Develop the Logic of Discovery in the Modern Age

In the seventeenth century it was widely held that the then current logic paradigm, Scholastic logic, was inadequate to the needs of the new science. Thus Bacon stated that Scholastic logic "is useless for the discovery of sciences;" it "is good rather for establishing and fixing errors (which are themselves based on vulgar notions) than for inquiring into truth; hence it is more harmful than useful" [11, I, p. 158]. Descartes stated that Scholastic logic "contributes nothing whatsoever to the knowledge of truth" [12, X, p. 406]. It "does not teach the method by which something has been discovered" (ibid., VII, p. 156). Therefore, it "is entirely useless for those who wish to investigate the truth of things" (ibid., X, p. 406).

The dissatisfaction with Scholastic logic gave rise to several attempts to develop logics of discovery alternative to Scholastic logic. The resulting logics of discovery, however, have serious limitations.

Thus, Bacon's logic is based on the use of tables involving a process of exclusion and rejection. But this process may either leave open too many possibilities, or none at all, therefore it may not lead to a necessary conclusion. To remedy this weakness, Bacon makes a plea for giving "the intellect permission" at this point to "try an interpretation of nature in the affirmative" [11, I, p. 261]. That is, Bacon makes a plea for giving the intellect permission to formulate a hypothesis, for example, a hypothesis about the nature of heat. But he does not really derive such hypothesis from an examination of the tables, rather, he borrows it from one of the views on heat which were discussed at the time. In fact, Bacon's method provides no means to formulate hypotheses.

On the other hand, Descartes' logic is ultimately based on intuition. According to Descartes, even deduction is based on intuition, because a deduction consists of a number of simple deductions, and "a simple deduction of one thing from another is performed by intuition" [12, X, p. 407]. Now, there are no rules for intuition. For such reason, Descartes states that his logic "cannot go so far as to teach us how to perform the actual operations of intuition and deduction, because these are the simplest and most primitive of all" (ibid., X, p. 372). Thus Descartes' method provides no means to formulate hypotheses.

That Bacon's and Descartes' logics have serious limitations does not mean, as Blanché states, that in their work "there is strictly nothing that deserves to be retained for the history of logic" [13, p. 174]. On the contrary, from their work we can learn that a logic of discovery can be based neither on a process of exclusion and rejection nor on intuition.

3 The Limitations of Mathematical Logic

The attempts to develop logics of discovery alternative to Scholastic logic fade in the nineteenth century and come to a definite end with Frege.

According to Frege, "the question of how we arrive at the content of a judgment should be kept distinct from the other question, Whence do we derive the justification for its assertion?" [14, p. 3]. Logic cannot be concerned with the former, the question of discovery, because it is a psychological question, "not a logical one" [15, p. 146]. It can be concerned only with the latter, the question of justification. For one cannot "count the grasping of the thought as knowledge, but only the recognition of its truth" (ibid., p. 267). In order to give a justification of a judgment, we must determine "upon what primitive laws it is based" [16, p. 235]. Then we must deduce the judgment from them. This will provide the required justification, since to deduce is "to make a judgment because we are cognisant of other truths as providing a justification for it" (ibid.). There are "laws governing this kind of justification," the laws of deduction, and "the goal of logic" is to study these laws, because they are the "laws of valid inference" (ibid.).

On this basis, mathematical logic has been developed as the logic of justification and as the study of the laws of deduction, avoiding the question of discovery. This attempt, however, has not been very successful. First, by Gödel's first incompleteness theorem, for any consistent sufficiently strong deductive theory T, there are sentences of T which are true but indemonstrable in T. In order to demonstrate them, one must discover new axioms. Therefore, mathematical logic cannot avoid the question of discovery. Secondly, by Gödel's second incompleteness theorem, for any consistent sufficiently strong deductive theory T, the sentence canonically expressing the consistency of T is not demonstrable by absolutely reliable means. Therefore, mathematical logic cannot be the logic of justification. Finally, by the strong incompleteness theorem for second-order logic, there is no set of deductive rules capable of deducing all second-order logical consequences of any given set of formulas. Therefore, mathematical logic is inadequate to the study of deduction. (For details, see [17], Introduction and Chap. 12).

Contrary to Frege, who considers the question of discovery a psychological one, Hilbert tries to trivialize it. On the one hand, he assumes that there is no question of discovering the axioms, because "the axioms can be taken quite arbitrarily" [18, p. 563]. They are only subject to the condition that "the application of the given axioms can never lead to contradictions" [19, p. 1093]. For "if the arbitrarily given axioms do not contradict one another, then they are true, and the things defined by the axioms exist. This for me is the criterion of truth and existence" [20, pp. 39-40]. On the other hand, Hilbert assumes that the question of discovering demonstrations of mathematical propositions from given axioms is a purely mechanical business. Since the axioms are arbitrarily given, this question—namely, the "decidability in a finite number of operations—is the best-known and the most discussed; for it goes to the essence of mathematical thought" [21, p. 1113].

Hilbert's attempt, however, fails because, on the one hand, by Gödel's second incompleteness theorem, it is impossible to show by absolutely reliable means that the application of the given axioms can never lead to contradictions. On the other hand, by the undecidability theorem, for any consistent sufficiently strong deductive theory, there is no mechanical procedure for deciding whether or not a mathematical proposition can be demonstrated from the axioms of the theory. Therefore, the question of the decidability of a mathematical question in a finite number of operations has a negative answer.

4 The Psychology of Discovery

Frege's view that logic cannot be concerned with the question of discovery because it is a psychological one, finds correspondence in Poincaré's proposal for a psychology of discovery.

According to Poincaré, "mathematical discovery" consists "in making new combinations" with concepts "that are already known" and in selecting "those that are useful" [22, pp. 50–51]. This is the action of the unconscious mind, which selects useful combinations on the basis of the "feeling of mathematical beauty" (ibid., p. 59). Once useful combinations have been selected, "it is necessary to verify them" (ibid., p. 56). This is the action of the conscious mind.

Poincaré's proposal for a psychology of discovery, however, is unconvincing.

First, if discovery consisted only in making new combinations with concepts that are already known, there should be some primitive concepts out of which all combinations of concepts would be made. Then, as Leibniz first pointed out, it would be possible to assign characters to primitive concepts, and form new characters for all other concepts, by means of combinations of such characters. The resulting characters would provide a universal language for mathematics, because it would be possible to express all mathematics concepts in terms of them. But this conflicts with Tarski's undefinability theorem, by which there cannot be a theory T capable of expressing all mathematical concepts, in particular, the concept of being a true sentence of T. Therefore, there cannot be a universal language for mathematics.

Secondly, the feeling of mathematical beauty, while sometimes useful, is generally unreliable as a means of selection of useful combinations. For example, the feeling of mathematical beauty led Galileo to stick to Copernicus' circular orbits for planets, which contrasted with observations, rejecting Kepler's elliptical orbits, which agreed with them. As another example, the feeling of mathematical beauty led Dirac to stick to his own version of quantum electrodynamics, which made predictions that were often infinite and hence unacceptable, rejecting renormalization, which led to accurate predictions.

5 The Need for a Rethinking of Logic

The failure of mathematical logic to provide a justification for truths already known, and the failure of the psychology of discovery to provide an account of the process of discovery, suggest that a rethinking of logic is necessary.

In particular, it is necessary to put a stop to the divorce of logic from method due to mathematical logic. Tarski states that there is "little rational justification for combining the discussion of logic and that of the methodology of empirical sciences" [23, p. xiii]. Consequently, as Aliseda points out, nowadays "logic (classical or otherwise) in philosophy of science is, to put it simply, out of fashion" [24, p. 21]. This contrasts with Aristotle's logic, which was developed as a tool for the method of science.

Because of the divorce of logic from method, mathematical logic has had little impact on scientific research. If logic is to play any significant role in science, an alternative logic paradigm is necessary. In particular, contrary to Scholastic logic and mathematical logic, in the alternative logic paradigm logic must be developed as a tool for the method of science and as the logic of discovery.

This is opposed by Laudan. He states that "the case has yet to be made that the rules governing the techniques whereby theories are invented (if such rules there be) are the sorts of things that philosophers should claim any interest in or competence at" [25, p. 182]. Therefore Laudan poses the challenge: "Why should the logic of discovery be revived?" (ibid.).

This question can be answered by saying that the logic of discovery should be revived, on the one hand, because, by Gödel's second incompleteness theorem, mathematical logic fails to be the logic of justification, and only reviving the logic of discovery, logic may continue to have an important role. On the other hand, scientists use heuristic tools in their work, and it may be useful to study such tools systematically in order to improve current heuristic tools and to develop new ones.

This means that logic must be developed as a tool for the method of science. But what is the method of science? Contemporary answers to this question involve two methods which, in their original form, were stated in antiquity: the analytic method and the analytic–synthetic method, where the latter includes the axiomatic method as its synthetic part.

6 The Analytic Method

The analytic method was first used by the mathematician Hippocrates of Chios and the physician Hippocrates of Cos and was first explicitly formulated by Plato. (On the original form of the method, see [17], Sects. 4.9, 4.13 and 4.18).

The analytic method is the method according to which, to solve a problem, one looks for some hypothesis from which a solution to the problem can be deduced. The hypothesis is obtained from the 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, that is, in accord with experience. But the hypothesis is in its turn a problem that must be solved, and is solved in the same way. That is, one looks 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 be of the same kind as the problem, and must be plausible. And so on, ad infinitum.

In the analytic method there are no principles, everything is a hypothesis. The problem and the other data already available are the only basis for solving the problem. A user of the analytic method is like Machado's walker: "Walker, your footsteps | are the road, and nothing more. | Walker, there is no road, | the road is made by walking" [26, p. 281].

The analytic method involves not only an upward movement, from problems to hypotheses, but also a downward movement, from hypotheses to problems, because one must examine the consequences of hypotheses in order to see whether they include a solution to the problem and are plausible.

The above statement of the analytic method is a revised version of Plato's original formulation. It differs from Plato's formulation in three respects.

1. Plato gives no indication as to how to find hypotheses to solve problems.
2. Plato only asks to examine the consequences of hypotheses in order "to see whether they are in accord, or are not in accord, with each other" (Plato, *Phaedo*, 101 d 4–5). This does not guarantee that they are in accord with experience.
3. Plato asks that knowledge be most certain and infallible, which leads him to conclude that "as long as we have the body and our soul is contaminated by such an

evil, we will never adequately gain the possession of what we desire, and that, we say, is truth" (ibid., 66 b 5–7). Thus we cannot reach knowledge during life.

The above statement of the analytic method is not subject to these limitations.

As to (1), it specifies that hypotheses are found by non-deductive rules.

As to (2), it asks that hypotheses be plausible, that is, in accord with experience.

As to (3), it does not ask that knowledge be most certain and infallible, but only plausible. This does not exclude that in the future new data may emerge with which the hypothesis is not in accord. (For more on the analytic method, see [17], Chap. 4).

7 The Analytic–Synthetic Method

The analytic–synthetic method was stated by Aristotle and is the basis of Aristotle's logic.

The analytic–synthetic method is the method according to which, to solve a problem of a given field, one must find premises for that field from which a solution to the problem can be deduced. According to Aristotle, the premises are obtained from the conclusion "either by syllogism or by induction" (Aristotle, *Topica*, Θ 1, 155 b 35–36). Moreover, the premises must be plausible, in the sense that they must be in accord with experience. If the premises thus obtained are not principles of the field in question, one must look for new premises from which the previous premises can be deduced. The new premises are obtained from the previous premises either by syllogism or by induction and must be plausible. And so on, until one arrives at premises which are principles of the field in question. Then the process terminates. This is analysis.

At this point one tries to see whether, inverting the order of the steps followed in analysis, one may obtain a deduction of the conclusion from the principles of the field in question, which must be known to be true by absolutely reliable means. This is synthesis. When synthesis is successful, this yields a solution to the problem.

8 Use of Non-deductive Rules

Both in the analytic and the analytic–synthetic method hypotheses are obtained by means of non-deductive rules. In the case of the analytic–synthetic method, this requires an explanation. As we have seen, according to Aristotle, in such method hypotheses are obtained from the conclusion either by syllogism or by induction. What does Aristotle mean by saying that they are obtained by syllogism?

Syllogism can be seen in a twofold manner: either as a means of obtaining conclusions from given premises, so as a means of justification, or as a means of obtaining premises for given conclusions, so as a means of discovery. According to a widespread view, for Aristotle syllogism is a means of justification, because he "shares with modern logicians the notion that central to the study of logic is examining the formal conditions for establishing knowledge of logical consequence" [27, p. 107]. This view is based on the first 26 Chapters of the first book of *Prior Analytics* where Aristotle describes the morphology of syllogism.

But in Chap. 27 Aristotle states: "Now it is time to tell how we will always find syllogisms on any given subject, and by what method we will find the premises about each thing. For surely one ought not only to investigate how syllogisms are constituted, but also to have the ability to produce them" (Aristotle, *Analytica Priora*, A 27, 43 a 20–24). In order to produce them, one must indicate "how to reach for premises concerning any problem proposed, in the case of any discipline whatever" (ibid., B 1, 53 a 1–2). That is, one must indicate, for any given conclusion, how to reach for premises from which that conclusion can be deduced.

From this it is clear that, for Aristotle, syllogism is primarily a means of discovery, specifically, a means for finding premises to solve problems. For this reason Aristotle says that, while "arguments are made from premises," the "things with which syllogisms are concerned are problems" (Aristotle, *Topica*, A 4, 101 b 15–16). Consistently with this view, in Chaps. 27-31 of the first book of *Prior Analytics* Aristotle describes a heuristic procedure for finding premises to solve problems. The medievals called this procedure *inventio medii* [discovery of the middle term] because it is essentially a procedure for finding the middle term of a syllogism, given the conclusion. (For a detailed description of this procedure, see [17], Sect. 7.4).

9 The Analytic–Synthetic Method and Modern Science

The originators of modern science adopted the analytic–synthetic method as the method of science. Contrary to a widespread opinion, the core of the Scientific Revolution of the seventeenth century was not a revolutionary change in the scientific method, but rather a change in the goal of science with respect to Aristotle. While, for Aristotle, the goal of science was to penetrate the true and intrinsic essence of natural substances, for Galileo it was only to know certain properties of natural substances, mathematical in character.

Indeed, Galileo famously stated: "Either, by speculating, we seek to penetrate the true and intrinsic essence of natural substances, or we content ourselves with coming to know some of their properties [*affezioni*]" [28, V, p. 187]. Trying to penetrate the essence of natural substances is "a not less impossible and vain undertaking with regard to the closest elemental substances than with the remotest celestial things" (ibid.). Therefore, we will content ourselves with coming to know "some properties of them," mathematical in character, "such as location, motion, shape, size, opacity, mutability, generation, and dissolution" (ibid., V, p. 188). While we cannot know the essence of natural substances, "we need not despair of our ability" to come to know such properties "even with respect to the remotest bodies, just as those close at hand" (ibid.).

But, while changing the goal of science, the originators of modern science adopted the analytic–synthetic method as the method of science. For example, Newton states: “In natural philosophy, the inquiry of difficult things by the method of analysis, ought ever to precede the method” of synthesis, or “composition” [29, p. 404]. Analysis “consists in making experiments and observations, and in drawing general conclusions from them by induction, and admitting of no objections against the conclusions, but such as are taken from experiments, or other certain truths” (ibid.). Synthesis or composition “consists in assuming the causes discovered, and established as principles, and by them explaining the phenomena proceeding from them, and proving the explanations” (ibid., p. 404–405). Newton’s own propositions “were invented by analysis,” then he composed, that is, wrote synthetically, what he had “invented by analysis” [30, p. 294].

10 Disadvantage of the Analytic–Synthetic Method

Despite its role as the method of modern science, the analytic–synthetic method has a serious disadvantage: it is incompatible with Gödel’s incompleteness theorems.

By Gödel’s first incompleteness theorem, for any consistent sufficiently strong principles of a given field, there are truths of that field which cannot be demonstrated from those principles. Their demonstration may require principles of other fields. Conversely, the analytic–synthetic method requires that every truth of a given field be deducible from principles of that field. Therefore, the analytic–synthetic method is incompatible with Gödel’s first incompleteness theorem.

On the other hand, by Gödel’s second incompleteness theorem, for any consistent, sufficiently strong principles, the principles cannot be known to be true by absolutely reliable means. Conversely, the analytic–synthetic method requires that principles be known to be true by absolutely reliable means. Therefore, the analytic–synthetic method is incompatible with Gödel’s second incompleteness theorem.

Since the analytic–synthetic method is incompatible with Gödel’s incompleteness theorems, the scientific method cannot be identified with it.

11 Advantages of the Analytic Method

Contrary to the analytic–synthetic method, the analytic method is compatible with Gödel’s incompleteness theorems. The latter even provide evidence for it.

The analytic method is compatible with Gödel’s first incompleteness theorem. In such method the solution to a problem is obtained from the problem, and possibly other data already available, by means of hypotheses which are not necessarily of the same field as the problem. Since Gödel’s first incompleteness theorem implies that solving a problem of a certain field may require hypotheses of other fields, Gödel’s result provides evidence for the analytic method.

The analytic method is also compatible with Gödel's second incompleteness theorem. In such method the hypotheses for the solution to a problem, being only plausible, are not absolutely certain. Since Gödel's second incompleteness theorem implies that no solution to a problem can be absolutely certain, Gödel's result provides evidence for the analytic method.

Not only the analytic method is compatible with Gödel's incompleteness theorems, but has several other advantages. (On the latter, see [17], Sects. 4.10, 5.17). In view of this, it seems reasonable to claim that the scientific method can be identified with the analytic method.

12 An Example

An example of use of the analytic method 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 $n > 2$.

Ribet showed that this problem could be solved using the hypothesis of Taniyama and Shimura, 'Every elliptic curve over the rational numbers is modular' (see [31]. But the hypothesis in question was in turn a problem that had to be solved. It was solved by Wiles using hypotheses from various mathematics fields. And so on.

13 The Paradox of Inference

In the analytic method, hypotheses are obtained by non-deductive rules, thus by rules which are not valid, that is, not truth preserving. That non-deductive rules cannot be valid follows from the so-called 'paradox of inference' (originally stated in [32], p.173).

According to the paradox of inference, if in an inference rule the conclusion is not contained in the premises, the rule cannot be valid; and if the conclusion does not possess novelty with respect to the premises, the rule cannot be ampliative; but the conclusion cannot be contained in the premises and also possess novelty with respect to them; therefore, an inference rule cannot be both valid and ampliative.

Since non-deductive rules are ampliative, from the paradox of inference it follows that they cannot be valid.

14 Objections Against the Use of Non-deductive Rules

Against the claim that the hypotheses for solving problems are obtained by the analytic method using non-deductive rules, some objections have been raised. I will consider two of them.

1. *Peirce's objection*. According to Peirce, the hypotheses for solving problems are obtained by abduction: "The surprising fact, C, is observed. But if A were true, C would be a matter of course. Hence, there is reason to suspect that A is true" [33, 5.189]. For example, Galileo observes the surprising fact C: 'Four bodies change their position around Jupiter'. But if A: 'Jupiter has satellites', were true, C would be a matter of course. 'Hence', Galileo concludes, 'there is reason to suspect that A is true'. According to Peirce, not only hypotheses for solving problems are obtained by abduction, but "all ideas of science come to it by way of abduction" (ibid., 5.145). Thus "abduction must cover all operations by which theories and conceptions are engendered" (ibid., 5.590).

This objection, however, is unjustified because abduction is of the form

$$\frac{C \quad A \rightarrow C}{A}$$

Thus the conclusion, A, already occurs as a part of the premise, $A \rightarrow C$, and hence is not really new. Peirce himself states that "A cannot be abductively inferred, or if you prefer the expression, cannot be abductively conjectured, until its entire content is already present in the premiss, 'If A were true, C would be a matter of course'" (ibid., 5.189). And he admits that "quite new conceptions cannot be obtained from abduction" (ibid., 5.190). This conflicts with his claim that abduction yields something new since it must cover all operations by which theories and conceptions are engendered.

Actually, a new hypothesis is not generated by abduction, but rather by the process that yields the premise, $A \rightarrow C$, thus a process prior to the application of abduction. As Franfurt points out, "clearly, if the new idea, or hypothesis, must appear in one of the premisses of the abduction, it cannot be the case that it originates as the conclusion of such an inference; it must have been invented before the conclusion was drawn" [34, p. 594].

Peirce claims that abduction is "what Aristotle meant by *apagogè*" [35, p. 140]. He assumes that Aristotle's *apagogè* is "the inference of the minor premiss" of a certain syllogism "from its other two propositions" [33, 7.251]. This, however, is unjustified because it contrasts with Aristotle's statement that "it is *apagogé* when it is clear that the first term belongs to the middle and unclear that the middle belongs to the third, though nevertheless more convincing than the conclusion" (Aristotle, *Analytica Priora*, B 25, 69 a 20–22). In order to support his claim, Peirce assumes that Aristotle's original text was corrupted, and the first editor filled the corrupted text "with the wrong word" (Peirce [35, p. 140]. But this is unconvincing, because Aristotle's text is perfectly intelligible as it stands. Peirce himself ends up acknowledging that his assumption is a "doubtful theory" [33, 8.209]. Rather than with *apagogè*, abduction can be compared with Aristotle's procedure of *inventio medii* mentioned in Sect. 8, but the comparison is unfavourable to abduction because, while in abduction the conclusion, A, already occurs as a part of the premise, $A \rightarrow C$, and hence is not really new, in *inventio medii* the middle term does not occur in the conclusion, and hence is new.

2. *Popper's objection*. According to Popper, hypotheses for solving problems cannot be obtained by non-deductive rules, in particular by induction, because we are not "justified in inferring universal statements from singular ones, no matter how numerous; for any conclusion drawn in this way may always turn out to be false" [36], p. 4). Generally, "there is no method of discovering a scientific theory" [37, p. 6]. Hypotheses are not obtained by a rational procedure but are rather "the result of an almost poetic intuition" [38, p. 192].

This objection, however, is unjustified because it is based on the following two tacit assumptions: (a) Knowledge must be known to be true; (b) Induction must provide a justification for the hypotheses it yields. Now, (a) is unwarranted because, by Gödel's second incompleteness theorem, it is impossible to know by absolutely reliable means if hypotheses are true. On the other hand, (b) is unwarranted because induction is only used to generate hypotheses, not to justify them, their justification depends on their accord with experience. By denying that induction may be used to obtain hypotheses, and generally that any method for obtaining hypotheses exists, Popper has no other option than saying that hypotheses are the result of an almost poetic intuition.

15 Scientific Knowledge and Certainty

If the analytic method is the method of science, then scientific knowledge is the result of solving problems by the analytic method. When solving problems is successful, this yields scientific knowledge.

Then scientific knowledge cannot be absolutely certain, because it is based on hypotheses that can only be plausible, and plausibility does not guarantee truth or certainty. On the other hand, plausibility is the best we can achieve, because there is no special source of knowledge capable of guaranteeing absolute truth or certainty. As Plato says, human beings can only "adopt the best and least refutable of human hypotheses, and embarking on it as on a raft, risk the dangers of crossing the sea of life" (Plato, *Phaedo*, 85 c 8–d 2).

That knowledge cannot be absolutely certain does not mean, as the sceptics claim, that knowledge is impossible, but only that absolutely certain knowledge is impossible.

16 Provisional Character of Solutions to Problems

As we have seen, in the analytic method solving a problem is a potentially infinite process, so no solution is final. At any stage, the inquiry may bring up new data incompatible with the hypothesis on which the solution is based. Then the hypothesis, while not being rejected outright, is put on a waiting list, subject to further inquiry. Even when the inquiry does not bring up incompatible data, the hypothesis remains a problem to be solved, and is solved by looking for another hypothesis, and so on. Thus any solution is provisional.

It might be objected that this does not account for the fact that mathematical results are final and forever. This objection, however, is unjustified because even solutions to mathematical problems are provisional. As Davis states, even in mathematics "a solved problem is still not completely solved but leads to new and profound challenges," and "discovering a sense in which" this is the case "is one important direction that mathematical research takes" [39, p. 177]. The analytic method suggests such a sense. A solved problem is still not completely solved since no hypothesis is absolutely justified. Any hypothesis which provides a solution to the problem is liable to be replaced with another hypothesis when new data emerge. Already Kant observed that "any answer given according to principles of experience always begets a new question which also requires an answer" [40, p. 103].

Quite generally, there is no final solution to problems. As Russell says, "final truth belongs to heaven, not to this world" [41, p. 3].

17 The Rules of Discovery

As we have seen, in the analytic method hypotheses are found by means of non-deductive rules. Of course, finding hypotheses is not a sufficient condition for discovery, because the latter requires hypotheses to be plausible. Nevertheless, finding hypotheses is a necessary condition for discovery and, in that sense, one may say that non-deductive rules are rules of discovery.

The latter, however, are not a closed set, given once for all, but rather an open set which can always be extended as research develops. Each such extension is a development of the analytic method, which grows as new non-deductive rules are added. In any case, the rules of discovery will at least include various kinds of induction and analogy, generalization, specialization, metaphor, metonymy, definition, and diagrams.

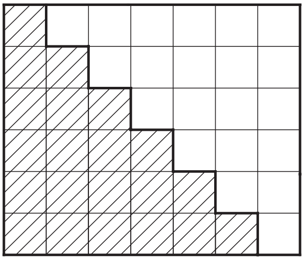
In what follows I will give an example of use of one of such rules in solving a problem. Such example is meant in the spirit of Pólya: "I cannot tell the true story how the discovery did happen, because nobody really knows that. Yet I shall try to make up a likely story how the discovery could have happened. I shall try to emphasize" the "inferences that led to it" [42, I, pp.vi–vii]. That nobody really knows the true story how the discovery did happen is due to the fact that discoverers generally do not reveal their way to discovery. They do so not because, as Descartes suggests, they conceal it "with a sort of pernicious cunning" [12, X, p. 376]. They do so rather because either they are not fully aware of how they arrived at their discoveries—awareness requires a good capacity for introspection—or feel uneasy to reveal that their way to discovery was not rigorously deductive.

The example concerns metaphor. Let $T \triangleright S$ stand for 'The elements of a domain, T, called the target domain, are to be considered as if they were elements of a domain, S, called the source domain'. Metaphor is an inference of the form: If $T \triangleright S$ and any arbitrary element of S has a certain property, then also any arbitrary element of T will have that property. So it is an inference by the rule:

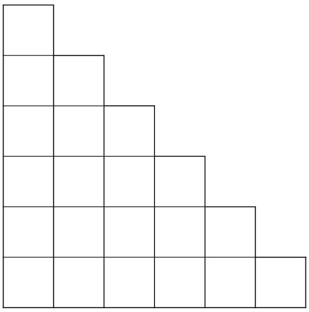
$$(\mathrm{MTA})\,\frac{T \rhd S \quad a \in T \quad \overset{[x\in S]}{A(x)}}{A(a)}.$$

Here $[x \in S]$ indicates that assumptions of the form $x \in S$ may be discharged by the rule.

An example of use of (MTA) is the Pythagoreans' discovery of the hypothesis $1 + 2 + \cdots + n = \frac{1}{2}n(n+1)$. From the following diagram, by (MTA), the Pythagoreans conclude the result.



The argument pattern is as follows. Let T be the domain of positive integers, and S be the domain of triangular figurate numbers, that is, geometric figures of the following form:



Let $T \rhd S$ and $n \in T$. Let $A(x)$: $1 + 2 + \cdots + x = \frac{1}{2}x(x+1)$. If $x \in S$ then $A(x)$. From this, by (MTA), it follows $A(n)$.

For other examples of rules of discovery and their use in solving problems, see [17], Chaps. 20 and 21.

18 Conclusion

The possibility of the logic of discovery is often denied by arguing that, since there are no inference rules "by which hypotheses or theories can be mechanically derived or inferred from empirical data," the "transition from data to theory requires creative imagination" [43, p. 15]. The "discovery of important, fruitful mathematical theorems, like the discovery of important, fruitful theories in empirical science, requires inventive ingenuity; it calls for imaginative, insightful guessing" (ibid., p. 17). That is, it calls for intuition.

This argument depends on the alternative: Either there are inference rules by which hypotheses can be mechanically inferred from empirical data, or the discovery of hypotheses calls for intuition. Since there are no inference rules by which hypotheses can be mechanically inferred from empirical data, the argument concludes that the discovery of hypotheses calls for intuition.

Now, the second horn of the alternative, that the discovery of hypotheses calls for intuition, is unsatisfactory because it amounts to admitting that, with respect to the discovery of new hypotheses or axioms, "there is no hope, there is, as it were, a leap in the dark, a bet at any new axiom. We are no longer in the domain of science but in that of poetry" [44, p. 169]. This would imply that science is an essentially irrational enterprise.

The first horn of the alternative, that there are inference rules by which hypotheses can be mechanically inferred from empirical data, corresponds to Bacon's assumption that, "in forming axioms, a form of induction, different from that hitherto in use, must be thought out" [11, I, p. 205]. Through it "the mind is from the very outset not left to itself, but constantly guided, so that everything proceeds, as it were, mechanically" (ibid., I, p. 152). Being mechanical, this form of induction is such as "not to leave much to the acuteness and strength of talents, but more or less to equalise talents and intellect (ibid., I, p. 172). Bacon's mechanical induction has had contemporary developments (see [45]). But Bacon's assumption is unwarranted. The purpose of the logic of discovery is not to dispense with the acuteness and strength of talents by use of mechanical rules, but rather to boost the acuteness and strength of talents by means of tools capable of guiding the mind, though not infallibly. Such a boost is provided by heuristic procedures which, rather than constantly guiding the mind so that everything proceeds mechanically, give the mind tools for finding hypotheses.

This shows that the alternative is unjustified. Between the determinism of mechanical rules and the inscrutability of intuition there is an intermediate region, inhabited by heuristic procedures. The latter consist of non-mechanical, non-deductive rules such as various kinds of induction and analogy, generalization, specialization, metaphor, metonymy, definition, and diagrams. These rules permit to find hypotheses by the analytic method without need to call for intuition.

Acknowledgments I am very grateful to Atocha Aliseda and Thomas Nickles for comments on an earlier draft of the manuscript.

References

1. Planck, M.: Scientific Autobiography and Other Papers. Williams & Norgate, London (1950)
2. Einstein, A.: Ideas and Opinions. Crown Publishers, New York (1954)
3. Einstein, A.: Induction and deduction in physics. In: Einstein, A. (ed.) Collected Papers, vol. 7, pp. 108–109. Princeton University Press, Princeton (2002)

4. Reichenbach, H.: The Rise of Scientific Philosophy. University of California Press, Berkeley (1951)
5. Reichenbach, H.: The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability. University of California Press, Berkeley (1949)
6. Novalis (von Hardenberg, Friedrich): Notes for a romantic encyclopedia. Das Allgemeine Brouillon. In: Wood DW (ed.) State University of New York Press, Albany (2007)
7. Cellucci, C.: Romanticism in mathematics. Newslett. HPM Group **64**, 17–19 (2007)
8. Lakatos, I.: Proofs and Refutations: The Logic of Mathematical Discovery. Cambridge University Press, Cambridge (1976)
9. Cellucci, C.: Top-down and bottom-up philosophy of mathematics. Found. Sci. **18**, 93–106 (2013)
10. Lakatos, I.: Philosophical papers. In: Worrall, J., Currie, G. (eds.) Cambridge University Press, Cambridge (1978)
11. Bacon, F.: Works. ed. In: Spedding, J., Ellis, R.L., Heath, D.D. (eds.) Frommann Holzboog, Stuttgart Bad Cannstatt (1961–1986)
12. Descartes, R.: *Œuvres*. In: Adam, C., Tannery, P (eds.). Vrin, Paris (1996)
13. Blanché, R.: La logique et son histoire, d'Aristote à Russell. Armand Colin, Paris (1970)
14. Frege, G.: The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number. Blackwell, Oxford (1959)
15. Frege, G.: Posthumous writings. In: Hermes, H., Kambartel, F., Kaulbach, F. (eds.). Blackwell, Oxford (1979)
16. Frege, G.: Collected papers on mathematics, logic, and philosophy. McGuinness, B. (ed.). Blackwell, Oxford
17. Cellucci, C.: Rethinking Logic. Logic in Relation to Mathematics, Evolution, and Method. Springer, Berlin (2013)
18. Hilbert, D.: Grundlagen der Geometrie. In: Hallett, M., Majer, U. (eds.) David Hilbert's Lectures on the Foundations of Geometry (1891–1902), pp. 540–606. Oxford University Press, Oxford (2004)
19. Hilbert, D.: On the concept of number. In: Ewald, W.B. (ed.) From Kant to Hilbert: A Source Book in the Foundations of Mathematics, II edn, pp. 1092–1095. Oxford University Press, Oxford (1996)
20. Hilbert, D.: Letter to Frege 29.12.1899. In: Gabriel, G., Hermes, H., Kambartel, F. (eds.) Gottlob Frege: Philosophical and Mathematical Correspondence, pp. 38–41. Blackwell, Oxford (1980)
21. Hilbert, D.: Axiomatic thought. In: Ewald, W.B. (ed.) From Kant to Hilbert: a source book in the foundations of mathematics, II edn, pp. 1107–1115. Oxford University Press, Oxford (1996)
22. Poincaré, H.: Science and Method. Nelson, London (1914)
23. Tarski, A.: Introduction to logic and to the methodology of deductive sciences. Tarski, J. (ed.). Oxford University Press, Oxford (1994)
24. Aliseda, A.: Abductive Reasoning: Logical Investigations into Discovery and Explanations. Springer, Berlin (2006)
25. Laudan, L.: Why Was the logic of discovery abandoned? In: Nickles, T. (ed.) Scientific Discovery, Logic, and Rationality, pp. 173–183. Reidel, Dordrecht (1980)
26. Machado, A.: Border of a dream: selected poems. In: Barnstone, W. (ed.). Copper Canyon Press, Port Townsend (2004)
27. Boger, G.: Aristotle's underlying logic. In: Gabbay, D.M., Woods, J. (eds.) Handbook of the History of Logic: Greek, Indian and Arabic logic, vol. 1, pp. 101–246. Elsevier, Amsterdam (2004)
28. Galileo, G.: Opere. In: Favaro, A. (ed.). Barbera, Florence (1968)
29. Newton, I.: Opticks, or a Treatise of the Reflections, Refractions, Inflections & Colours of Light. Dover, Mineola (1952)
30. Cohen, I.B.: Introduction to Newton's 'Principia'. Cambridge University Press, Cambridge (1971)

31. 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)
32. Morris, R.C., Nagel, E.: An Introduction to Logic and Scientific Method. Routledge, London (1964)
33. Peirce, C.S.: Collected papers. In: Hartshorne, C., Weiss, P., Burks, A.W. (eds.). Cambridge University Press, Cambridge (1931–1958)
34. Frankfurt, H.G.: Peirce's notion of abduction. J. Philos. **55**, 593–597 (1958)
35. Peirce, C.S.: Reasoning and the logic of things. In: Ketner KL (ed.). Harvard University Press, Cambridge (1992)
36. Popper, K.: The Logic of Scientific Discovery. Routledge, London (2005)
37. Popper, K.: Realism and the aim of science. In: Bartley WW (ed.), III. Routledge, London (2000)
38. Popper, K.: Conjectures and Refutations: The Growth of Scientific Knowledge. Routledge, London (1974)
39. Davis, P.J.: Mathematics and Common Sense: A Case of Creative Tension. A K Peters, Natick (2006)
40. Kant, I.: Prolegomena to Any Future Metaphysics. Cambridge University Press, Cambridge (2004)
41. Russell, B.: An Outline of Philosophy. Routledge, London (1995)
42. Pólya, G.: Mathematics and Plausible Reasoning. Princeton University Press, Princeton (1954)
43. Hempel, C.G.: Philosophy of Natural Science. Prentice Hall, Upper Saddle River (1966)
44. Girard, J.Y.: Le champ du signe ou la faillite du réductionnisme. In: Nagel, E., Newman, J.R., Gödel, K., Girard, J.Y. (eds.) Le théorème de Gödel, pp. 147–171. Éditions du Seuil, Paris (1989)
45. Gillies, D.: Artificial Intelligence and Scientific Method. Oxford University Press, Oxford (1996)

Are Heuristics Knowledge–Enhancing? Abduction, Models, and Fictions in Science

Lorenzo Magnani

> *It is true that the different elements of the hypothesis were in our minds before; but it is the idea of putting together what we had never before dreamed of putting together which flashes the new suggestion before our contemplation.*
>
> Charles Sanders Peirce, *Collected Papers*, 5.181

Abstract In my opinion, it is only in the framework of a study concerning abductive inference that we can correctly and usefully grasp the cognitive status of heuristics. To this aim, it is useful to see heuristics in the perspective of the so-called fill-up and cutdown problems, which characterize abductive cognition. Abduction is a procedure in which something that lacks classical explanatory epistemic virtue can be accepted because it has virtue of another kind: [9] contend (GW-Model) that abduction presents an *ignorance-preserving* or (ignorance-mitigating) character. The question is: are abductive heuristic strategies always ignorance preserving? To better reframe the cognitive status of heuristics I will take advantage of my *eco-cognitive model* (EC-model) of abduction. I contend that, through abductive heuristics, knowledge can be enhanced, even when abduction is not considered an inference to the best explanation in the classical sense of the expression, that is an inference necessarily characterized by an empirical evaluation phase, or inductive phase, as Peirce called it. Hintikka maintains, implicity agreeing with the perspective on abduction as ignorance-preserving, that the true justification of a rule of abductive inference is a strategic one, but this strategic justification *does not warrant* any specific step of the whole process. I argue, taking advantage of a distinction between static and dynamic aspects of scientific inquiry, that this does mean that every abductive guess heuristically reached is damned to be ignorance-preserving if evidentially inert. When Hintikka contends that the abductive steps which lead to intermediate models *cannot* have "warrants" at the level of strategic justification, and also at the level of non strategic justification, in my perspective we can relieve ourselves of this burden

L. Magnani (✉)
Department of Humanities, Philosophy Section, and Computational Philosophy Laboratory, University of Pavia, Pavia, Italy
e-mail: lmagnani@unipv.it

E. Ippoliti (ed.), *Heuristic Reasoning*, Studies in Applied Philosophy, Epistemology and Rational Ethics 16, DOI 10.1007/978-3-319-09159-4_3

of epistemic sufferance just acknowledging we are dealing with *creative* models. If we only see models in empirical science in the light of the future achieved empirical success, we obviously see them just as *provisional* guesses, *devoid* of justification and still and intrinsically looking for it. On the contrary, they are occasionally justified by themselves—abductively—just because creative, and so *constitutive* of a fruitful epistemic "heuristic cognitive travel". Finally, I will illustrate that also in deduction the presence of abductive heuristic events coincides with their knowledge-enhancing character: here too these strategic aspects reflect the pure—productive—conjectural element of abductive inference and its capacity to guessing right.

Keywords Abduction · Heuristics · Models · Fictions · Creativity

1 Is Abduction Ignorance-Preserving?

The general form of an abductive inference can be symbolically rendered as follows. Let α be a proposition with respect to which you have an ignorance problem. Putting T for the agent's epistemic target with respect to the proposition α at any given time, K for his knowledge-base at that time, K^* for an immediate accessible successor-base of K that lies within the agent's means to produce in a timely way,[1] R as the attainment relation for T, $\rightsquigarrow$, as the *subjunctive* conditional relation, H as the agent's hypothesis, $K(H)$ as the revision of K upon the addition of H, $C(H)$ denotes the conjecture of H and H^c its activation. The general structure of abduction can be illustrated as follows (GW-schema)[2]:

1. $T!\alpha$	[Setting of T as an epistemic target with respect to a proposition α]
2. $\neg(R(K,T)$	[fact]
3. $\neg(R(K^*,T)$	[fact]
4. $H \notin K$	[fact]
5. $H \notin K^*$	[fact]
6. $\neg R(H,T)$	[fact]
7. $\neg R(K(H),T)$	[fact]
8. If $H \rightsquigarrow R(K(H),T)$	[fact]
9. H meets further conditions $S_1, \ldots S_n$	[fact]
10. Therefore, $C(H)$	[sub-conclusion, 1-9]
11. Therefore, H^c	[conclusion, 1-10]

[1] K^* is an accessible successor of K to the degree that an agent has the know-how to construct it in a timely way; i.e., in ways that are of service in the attainment of targets linked to K. For example if I want to know how to spell "accommodate", and have forgotten, then my target can't be hit on the basis of K, what I now know. But I might go to my study and consult the dictionary. This is K^*. It solves a problem originally linked to K.

[2] That is Gabbay and Woods Schema.

It is easy to see that the distinctive epistemic feature of abduction is captured by the schema. It is a given that H is not in the agent's knowledge-set. Nor is it in its immediate successor. Since H is not in K, then the revision of K by H is not a knowledge-successor set to K. Even so, $H \rightsquigarrow (K(H), T)$. So we have an ignorance-preservation, as required [44, Chap. 10].

[*Note*: Basically, line 9. indicates that H has no more plausible or relevant rival constituting a greater degree of subjunctive attainment. Characterizing the S_i is the most difficult problem for abductive cognition, given the fact that in general there are many possible candidate hypotheses. It involves for instance the *consistency* and *minimality* constraints.[3] These constraints correspond to the lines 4 and 5 of the standard AKM schema of abduction,[4] which is illustrated as follows:

1. E
2. $K \not\looparrowright E$
3. $H \not\looparrowright E$
4. $K(H)$ is consistent
5. $K(H)$ is minimal
6. $K(H) \looparrowright E$
7. Therefore, H.

[9, pp. 48–49].

where of course the conclusion operator $\looparrowright$ cannot be classically interpreted].[5]

Finally, in the GW-schema $C(H)$ is read "It is justified (or reasonable) to conjecture that H" and H^c is its activation, as the basis for *planned* "actions". Without this last step, and coherently with the above schema, abduction has to be considered an *ignorance-preserving* inference.

In sum, in the GW-schema T cannot be attained on the basis of K. Neither can it be attained on the basis of any successor K^* of K that the agent knows then and

[3] I have shown [23, Chap. 2, Sect. 2.3.1] that, in the case of inner processes in organic agents, this sub-process—here explicitly modeled thanks to a formal schema—is considerably implicit, and so also linked to unconscious ways of inferring, or even, in Peircean terms, to the activity of the instinct [32, 8.223] and of what Galileo called the *lume naturale* [32, 6.477], that is the innate fair for guessing right. This and other cognitive aspects can be better illustrated thanks to the alternative eco-cognitive model (EC-Model) of abduction I have illustrated in my book [23].

[4] The classical schematic representation of abduction is expressed by what [9] call AKM-schema, which is contrasted to their own (GW-schema), which I am just explaining in this section. For A they refer to Aliseda [1, 2], for K to Kowalski [16], Kuipers [17], and Kakas et al. [15], for M to Magnani [21] and Meheus et al. [28]. A detailed illustration of the AKM schema is given in [23, Chap. 2, Sect 2.1.3].

[5] The target has to be an explanation and $K(H)$ bears R^{pres} [that is the relation of presumptive attainment] to T only if there is a proposition V and a consequence relation $\looparrowright$ such that $K(H) \looparrowright V$, where V represents a *payoff proposition* for T. In turn, in this schema explanations are interpreted in consequentialist terms. If E is an explanans and E' an explanandum the first explains the second only if (some authors further contend if and only if) the first implies the second. It is obvious to add that the AKM schema embeds a D-N (deductive-nomological) interpretation of explanation, as I have already stressed in [21, p. 39]. .

there how to construct. H is not in K: H is a hypothesis that when reconciled to K produces an updated $K(H)$. H is such that if it were true, then $K(H)$ would attain T. The problem is that H is *only hypothesized*, so that the truth is not assured. Accordingly Gabbay and Woods contend that $K(H)$ *presumptively* attains T. That is, having hypothesized that H, the agent just "presumes" that his target is now attained. Given the fact that presumptive attainment is not attainment, the agent's abduction must be considered as preserving the ignorance that already gave rise to her (or its, in the case for example of a machine) initial ignorance-problem. Accordingly, abduction does not have to be considered the "solution" of an ignorance problem, but rather a response to it, in which the agent reaches presumptive attainment rather than actual attainment. $C(H)$ expresses the conclusion that it follows from the facts of the schema that H is a worthy object of conjecture. It is important to note that in order to solve a problem it is not necessary that an agent actually conjectures a hypothesis, but it is necessary that she states that the hypothesis is *worthy of conjecture*.

Finally, considering H justified to conjecture is not equivalent to considering it justified to accept/activate it and eventually to send H to experimental trial. H^c denotes the *decision* to release H for further promissory work in the domain of enquiry in which the original ignorance-problem arose, that is the activation of H as a positive *cognitive* basis for action. Woods usefully observes:

> There are lots of cases in which abduction stops at line 10, that is, with the conjecture of the hypothesis in question but not its activation. When this happens, the reasoning that generates the conjecture does not constitute a positive basis for new action, that is, for acting *on* that hypothesis. Call these abductions *partial* as opposed to full. Peirce has drawn our attention to an important subclass of partial abductions. These are cases in which the conjecture of H is followed by a decision to submit it to experimental test. Now, to be sure, doing this is an action. It is an action *involving* H but it is not a case of acting *on* it. In a full abduction, H is activated by being released for inferential work in the domain of enquiry within which the ignorance-problem arose in the first place. In the Peircean cases, what counts is that H is withheld from such work. Of course, if H goes on to test favourably, it may then be released for subsequent inferential engagement [42].

We have to remember that this process of evaluation and so of activation of the hypothesis, is not abductive, but inductive, as Peirce contended. Woods adds: "Now it is quite true that epistemologists of a certain risk-averse bent might be drawn to the admonition that partial abduction is as good as abduction ever gets and that complete abduction, inference-activation and all, is a mistake that leaves any action prompted by it without an adequate rational grounding. This is not an unserious objection, but I have no time to give it its due here. Suffice it to say that there are real-life contexts of reasoning in which such conservatism is given short shrift, in fact is ignored altogether. One of these contexts is the criminal trial at common law" [42].

In the framework of the GW-schema it cannot be said that testability is intrinsic to abduction, such as it is instead maintained in the case of some passages of Peirce's writings.[6] This activity of testing, I repeat, which in turn involves degrees

[6] When abduction stops at line 10. (cf. the GW-schema), the agent is not prepared to accept $K(H)$, because of supposed adverse consequences.

of risk proportioned to the strength of the conjecture, is strictly cognitive/epistemic and inductive in itself, for example an experimental test, and it is an intermediate step to release the abduced hypothesis for inferential work in the domain of enquiry within which the ignorance-problem arose in the first place.

Through abduction the basic ignorance—that does not have to be considered total "ignorance"—is neither solved nor left intact: it is an ignorance-preserving accommodation of the problem at hand, which "mitigates" the initial cognitive "irritation" (Peirce says "the irritation of doubt").[7] As I have already stressed, further action can be triggered—in a defeasible way—either to find further abductions or to "solve" the ignorance problem, possibly leading to what the "received view" has called the *inference to the best explanation* (IBE).

It is clear that in the framework of the GW-schema the inference to the best explanation—if considered as a truth conferring achievement justified by the empirical approval—cannot be a case of abduction, because abductive inference is constitutively ignorance-preserving. In this perspective the inference to the best explanation involves the generalizing[8] and evaluating role of *induction*. Of course it can be said that the requests of originary thinking are related to the depth of the abducer's ignorance.

We can usefully see selective and creative abduction[9] as often formed by the application of "heuristic procedures" that involve all kinds of good and bad inferential moves, and not only the mechanical application of rules. It is by means of these heuristic procedures that the acquisition of *new* truths is guaranteed. Also Peirce's view on creative abduction as a kind of heuristic inference—cf. the epigraph placed at the beginning of this article—seems to stress the strategic component of reasoning. Given the fact a considerable part of abductive reasoning can be seen as performed through heuristics, have they to be considered ignorance-preserving? To answer this question I will provide in the following sections a detailed

[7] "The action of thought is excited by the irritation of doubt, and ceases when belief is attained; so that the production of belief is the sole function of thought" [33, p. 261].

[8] By illustrating abductive/inductive reasoning of preservice elementary majors on patterns that consist of figural and numerical cues in learning elementary mathematics Rivera and Becker monitor the subsequent role of induction. In performing the abductive task to the general form/hypothesis the subjects referred to the fact they immediately saw a relationship among the drawn cues in terms of relational similarity "[…] within classes in which the focus was *not* on the individual clues in a class per se but on a possible invariant relational structure that was perceived between and, thus, projected onto the cues" [37, p. 151]. Through the follow-up inductive stage of generalizations the subjects tested the hypotheses just examining *extensions* (new particular cases beyond what was available at the beginning of the reasoning process). This process was also able to show subjects's disconfirmation capacities: they acknowledged their mistakes in generating a bad induction, which had to be abandoned, in so far as they were checked as insufficient in fully capturing in symbolic terms a general attribute that would yield the total number of toothpicks in new generated cues.

[9] In diagnostic reasoning, for example, abduction is merely seen as an activity of "selecting" from an encyclopedia of pre-stored hypotheses. Creative abduction instead generates new hypotheses. I have proposed the dichotomic distinction between selective and creative abduction in [21].

analysis of both cutdown and fill-up problems in abduction and of the role of models and fictions in science, arguing for a possible knowledge-enhancing character of heuristics.

2 Heuristics in the Perspective of Cutdown and Fill-Up Problems

2.1 Cutdown and Fill-Up Problems in the EC-Model of Abduction

In my opinion, it is only in the framework of abductive inference that we can correctly and usefully grasp the cognitive problem of heuristics. To this aim, it is useful to see heuristics in the perspective of the so-called fill-up and cutdown problems (see below in this subsection), which characterize abductive cognition. From a general philosophical perspective (with, and beyond, Peirce) the condition 9. (cf. the GW-schema) is, as Woods himself admits "more a hand-wave than a real condition. Of course the devil is in the details. [...] I myself I am not sure" [45, p. 242]. Obviously consistency and minimality constraints were emphasized in the "received view" on abduction established by many classical logical accounts, more oriented to illustrate selective abduction [21]—for example in diagnostic reasoning, where abduction is merely seen as an activity of "selecting" from an encyclopedia of pre-stored hypotheses—than to analyze *creative* abduction (abduction that generates new hypotheses).[10]

For example, to stress the puzzling status of the consistency requirement, it is here sufficient to note that Paul Feyerabend, in *Against Method* [7], correctly attributes a great importance to the role of contradiction in generating hypotheses, also against the role of similarity, and so implicity celebrates the value of creative abductive cognition. Speaking of induction and not of abduction (this concept was relatively unknown at the level of the international philosophical community at that time), he establishes a new "counterrule". This is the opposite of the neopositivistic one that it is "experience" (or "experimental results") which constitutes the most important part of our scientific empirical theories, a rule that formed the core of the so-called "received view" in philosophy of science (where inductive generalization, confirmation, and corroboration play a central role). The counterrule "[...] advises us to introduce and elaborate hypotheses which are inconsistent with well-established theories and/or well-established facts. It advises us to proceed counterinductively" [7, p. 20]. Counterinduction is seen more reasonable than induction, because appropriate to the needs of creative reasoning in science: "[...] we need a dream-world in order to discover the features of the real world we think we inhabit" (p. 29). We know that counterinduction, that is the act of introducing, inventing, and generating new inconsistencies and anomalies, together with new points of view incommensurable with the old ones, is congruous with the aim

[10] See the previous footnote.

of inventing "alternatives" (Feyerabend contends that "proliferation of theories is beneficial for science"), and very important in all kinds of creative reasoning.

Since for many abduction problems there are—usually—many guessed hypotheses, the abducer needs reduce this space to one: this means that the abducer has to produce the best choice among the members of the available group: "It is extremely difficult to see how this is done, both formally and empirically. Clause (9) [in the GW-model] is a place-holder for two problems, not one. There is the problem of finding criteria for hypothesis *selection*. But there is the prior problem of specifying the conditions for *thinking up* possible candidates for selection. The first is a 'cutdown' problem. The second is a 'fill-up problem'; and with the latter comes the received view that it is not a problem for logic" [45, p. 243] emphasis added).

Here we touch the core of the ambiguity of the ignorance-preserving character of abduction. Why?

- Because the cognitive processes of generation (fill-up) and of selection (cutdown) can both be sufficient—even in absence of the standard inductive evaluation phase—to *activate* and accept [clause (11) of the GW-schema above] an abductive hypothesis, and so to reach cognitive results relevant to the context (often endowed with a knowledge-enhancing outcome, as I have illustrated in [26]. In these cases instrumental aspects (which simply enable one's target to be hit) often favor both abductive generation and abductive choice, and they are not necessarily intertwined with plausibilistic concerns, such as consistency and minimality.

In these special cases the best choice is—often thanks to the exploitation of heuristics—immediately reached without the help of an experimental trial (which fundamentally characterizes the received view of abduction in terms of the so-called "inference to the best explanation").

Let us recall some basic information concerning the received view on strategic rules. Hintikka thinks that "strategic rules" (contrasted with definitory rules) are smart rules, even if they fail in individual cases, and show a propensity for cognitive success. In the case of abduction, they tacitly fulfil the ignorance condition I have illustrated in Sect. 1, thus abduction would aim at neither truth-preservation not probability-enhancement, as Peirce maintained. In a strict sense, Hintikka's heuristic rules are strategic rules, even if merely tentative and partial [11, p. 69]. Moreover, Hintikka's definitory rules are recursive but in several important cases strategic rules are not: therefore, playing a game strategically requires some kind of creativity.

I contended—few lines above—that special cases of cognitive processes of generation (fill-up) and of selection (cutdown) are sufficient to reach the acceptation of a hypothesis: even in absence of the standard inductive evaluation phase the best choice is immediately reached without the help of an experimental trial (which fundamentally characterizes the received view of abduction in terms of the so-called "inference to the best explanation"). The best choice is often reached through heuristics, that are not, in these cases, ignorance preserving, but instead knowledge-enhancing. Furthermore, we have to strongly note that the generation process alone can be still seen sufficient considering the case of human *perception*, where the hypothesis generated is immediate and unique, even if no heuristics appear to be involved. Indeed, perception is considered by Peirce, as an "abductive" fast and uncontrolled (and so automatic) knowledge-production procedure. Perception, in this philosophical perspective, is a vehicle for the instantaneous retrieval of knowledge that was previously structured in our mind through more structured inferential processes. Peirce says: "Abductive inference shades into perceptual judgment without any sharp line of demarcation between them" [34, p. 304]. By perception, knowledge constructions are so instantly reorganized that they become habitual and diffuse and do not need any further testing: "[…] a fully accepted, simple, and interesting inference tends to obliterate all recognition of the uninteresting and complex premises from which it was derived" [32, 7.37].[11]

My abrupt reference to perception as a case of abduction (in this case I strictly follow Peirce) does not have to surprise the reader. Indeed, at the of center of my perspective on cognition is the emphasis on the "practical agent", of the individual agent operating "on the ground", that is, in the circumstances of real life. In all its contexts, from the most abstractly logical and mathematical to the most roughly empirical, I always emphasize the cognitive nature of abduction. Reasoning is something performed by cognitive systems. At a certain level of abstraction and as a first approximation, a cognitive system is a triple (A, T, R), in which A is an *agent*, T is a *cognitive target* of the agent, and R relates to the *cognitive resources* on which the agent can count in the course of trying to meet the target-information, time and computational capacity, to name the three most important. My agents are also *embodied distributed cognitive systems*: cognition is embodied and the interactions between brains, bodies, and external environment are its central aspects. Cognition is occurring taking advantage of a constant exchange of information in a complex distributed system that crosses the boundary between humans, artifacts, and the surrounding environment, where also instinctual and unconscious abilities play an important role. This interplay is especially manifest and clear in various aspects of abductive cognition.[12]

It is in this perspective that we can appropriately consider perceptual abduction as a fast and uncontrolled knowledge production, that operates for the most part automatically and out of sight, so to speak. This means that—at least in this light—GW-schema is not canonical for abduction. The schema illustrates what I call "sentential abduction" [23, chapter one], that is, abduction rendered by symbols carrying propositional content. It is hard to encompass in this model cases of abductive cognition such as perception or the generation of models in scientific

[11] A relatively recent cognitive research related to artificial intelligence presents a formal theory of robot perception as a form of abduction, so reclaiming the rational relevance of the speculative anticipation furnished by Peirce, cf. [38].

[12] It is interesting to note that recent research on Model Checking in the area of AST (Automated Software Testing) takes advantage of this eco-cognitive perspective, involving the manipulative character of model-based abduction in the practice of adapting, abstracting, and refining models that do not provide successful predictions. Cf. [3].

discovery (cf. Sect. 3.2). My perspective adopts the wide Peircean philosophical framework, which approaches "inference" *semiotically* (and not simply "*logically*"): Peirce distinctly says that all inference is a form of sign activity, where the word sign includes "feeling, image, conception, and other representation" [32, 5.283]. It is clear that this semiotic view is considerably compatible with my perspective on cognitive systems as embodied and distributed systems: the GW-Schema is instead only devoted to illustrate, even if in a very efficacious way, a subset of the cognitive systems abductive activities, the ones that are performed taking advantage of explicit propositional contents. Woods seems to share this conclusion: "[…] the GW-model helps get us started in thinking about abduction, but it is nowhere close, at any level of abstraction, to running the whole show. It does a good job in modelling the ignorance-preserving character of abduction; but, since it leaves the S_i of the schema's clause (T) unspecified, it makes little contribution to the fill-up problem" [45, p. 244].

In a wide eco-cognitive perspective the cutdown and fill-up problems in abductive cognition appear to be spectacularly *contextual*.[13] I lack the space to give this issue appropriate explanation but it suffices for the purpose of this study—which instead aims at revisiting the concept of heuristics—to remember that, for example, one thing is to abduce a model or a concept at the various levels of scientific cognitive activities, where the aim of reaching rational knowledge dominates, another thing is to abduce a hypothesis in literature (a fictional character for example), or in moral reasoning (the adoption/acceptation of a hypothetical judgment as a trigger for moral actions). However, in all these cases abductive hypotheses which are evidentially inert are accepted and activated as a basis for action, even if of different kind.

The backbone of this approach can be found in the manifesto of my EC-model of abduction in [23]. It might seem awkward to speak of "abduction of a hypothesis in literature," but one of the fascinating aspects of abduction is that not only it can warrant for scientific discovery, but for other kinds of creativity as well. We must not necessarily see abduction as a *problem solving device* that sets off in response to a cognitive irritation/doubt: conversely, it could be supposed that esthetic abductions (referring to creativity in art, literature, music, etc.) arise in response to some kind of esthetic irritation that the author (sometimes a *genius*) perceives in herself or in the public. Furthermore, not only esthetic abductions are free from empirical constraints in order to become the "best" choice: as I am showing throughout this paper, many forms of abductive hypotheses in traditionally-perceived-as-rational domains (such as the setting of initial conditions, or axioms, in physics or mathematics) are relatively free from the need of an empirical assessment. The same could be said of moral judgements: they are eco-cognitive abductions, inferred upon a range of internal and external cues and, as soon as the judgment hypothesis has been abduced, it immediately becomes prescriptive and

[13] Some acknowledgment of the general contextual character of these kinds of criteria, and a good illustration of the role of coherence, unification, explanatory depth, simplicity, and empirical adequacy in the current literature on scientific abductive best explanation, is given in [20].

"true," informing the agent's behavior as such. Assessing that there is a common ground in all of these works of what could be broadly defined as "creativity" does not imply that all of these forms of selective or creative abduction are the same, contrarily it should spark the need for firm and sensible categorization: otherwise it would be like saying that to construct a doll, a machine-gun and a nuclear reactor are all the same thing because we use our hands in order to do so!

2.2 Do Heuristic Strategies Justify Abduction? Oracles and Ignorance Preservation

I have just illustrated the main theoretical tools we need to reframe heuristics in an eco-cognitive perspective on abduction. Indeed, from an eco-cognitive point of view, in more hybrid and multimodal [23, Chap. 4] (not merely inner) abductive processes, such as in the case of manipulative abduction,[14] the *assessment/acceptation* of a hypothesis is reached—and constrained—taking advantage of the gradual—strategic—acquisition of consecutive external information with respect to future interrogation and control, and not necessarily thanks to a final and actual experimental test, in the classical sense of empirical science.

Hintikka implicitly acknowledges the multimodality and hybridity of what I call *selective abduction* when, taking advantage of the intellectual atmosphere of his Socratic interrogative epistemology, observes that "[…] abduction as a method of guessing is based on the variety of different possible sources of answers. Such 'informants' must include not only testimony, observation, and experiments, but the inquirer's memory and background knowledge" [12, p. 56]. Moreover, Hintikka further notes that also "creative abduction", generated by a kind of *oracle*, is often needed: "But what can an inquirer do when all such sources fail to provide an answer to a question? Obviously the best the inquirer can do is make an informed guess. For the purposes of a general theory of inquiry, what Peirce calls 'intelligent guessing' must therefore be recognized as one of the many possible 'oracles', alias sources of answers. Peirce may very well have been more realistic than I have so far been in emphasizing the importance of this particular 'oracle' in actual human inquiry" (ibid.).

In summary, at least four kinds of actions can be involved in the manipulative abductive processes (and we would have to also take into account the motoric aspect (i) of inner "thoughts" too). In the eco-cognitive interplay of abduction the cognitive agent further triggers internal *thoughts* "while" modifying the

[14] The concept of *manipulative abduction*—which also takes into account the external dimension of abductive reasoning in an eco-cognitive perspective—captures a large part of scientific thinking where the role of action and of external models (for example diagrams) and devices is central, and where the features of this action are implicit and hard to be elicited. Action can provide otherwise unavailable information that enables the agent to solve problems by starting and by performing a suitable abductive process of generation and/or selection of hypotheses. Manipulative abduction happens when we are thinking through doing and not only, in a pragmatic sense, about doing [23, Chap. 1]. Cf. also below, Sect. 3.3.

environment and so (ii) acting on it (thinking through doing). In this case the "motor actions" directed to the environment have to be intended as part and parcel of the whole embodied abductive inference, and so have to be distinguished from the final (iii) "actions" as a possible consequence of the reached abductive result.

In this perspective the proper experimental test involved in the Peircean evaluation phase, which for many researchers reflects in the most acceptable way the idea of abduction as inference to the best explanation, just constitutes a *special* subclass of the process of the adoption of the abductive hypothesis—the one which involves a terminal kind (iv) of actions (experimental tests), and should be considered ancillary to the nature of abductive cognition, and inductive in its essence. We have indeed to remark again that in Peirce's mature perspective on abduction as embedded in a cycle of reasoning, induction just plays an evaluative role. Hintikka usefully notes, and I agree with him, that Peirce was right in denying the role of "naked" induction in forming new hypotheses:

> Many philosophers would probably bracket abductive inference with inductive inference. Some would even think of all ampliative inference as being, at bottom, inductive. In this matter, however, Peirce is one hundred percent right in denying the role of naked induction in forming new hypotheses. […] It might seem that the critical and evaluative aspect of inquiry that Peirce called inductive still remains essentially different from the deductive and abductive aspects. A common way of thinking equates all ampliative inferences with inductive ones. Peirce was right in challenging this dichotomy. Rightly understood, the ampliative versus non-ampliative contrast becomes a distinction between interrogative (ampliative) and deductive steps of argument. As in Peirce, we also need over and above these two also the kind of reasoning that is involved in testing the propositions obtained as answers to questions. I do not think that it is instructive to call such reasoning inductive, but this is a merely terminological matter [12, pp. 52 and 55].

In absence of empirical evaluation, can we attribute the *pure* abductive inclination to produce right guesses indicated by Peirce, conductive to the acquisition of truth, to the *reliability* of the process? Yes, we can, but only if we take into account the following warning, still illustrated by Hintikka, who stresses the importance of *strategic/heuristic aspects*, arguing for a fundamental role of them as the warrant and justification of abductive inference: "Many contemporary philosophers will assimilate this kind of justification to what is called a reliabilist one. Such reliabilist views are said to go back to Frank Ramsey, who said that 'a belief was knowledge if it is (1) true, (2) certain, (3) *obtained by a reliable process*' (emphasis added). Unfortunately for reliabilists, such characterizations are subject to the ambiguity that was pointed out earlier. By a reliable process one can mean either a process in which each step is conducive to acquiring and/or maintaining truth or closeness to truth, or one that as a whole is apt to lead the inquirer to truth. Unfortunately, most reliabilists unerringly choose the wrong interpretation—namely, the first one. As was pointed out earlier, the true justification of a rule of abductive inference is a strategic one" [12, p. 57]. The important thing is to stress that this strategic justification *does not warrant* any specific step of the whole process. Let us remember that abduction certainly provides new information into an argument, but this is not necessarily a true information, because it is not implied by what it is already known or accepted but it is constitutively hypothetical—that is, ignorance-preservation is constitutive, from the general logico-philosophical point of view, and Hintikka is in tune with this assumption.

3 Knowledge-Enhancing Abductive Heuristics

3.1 Is Abduction Knowledge-Enhancing?

Even if abduction, in the perspective of the formal GW-model above, is ignorance-preserving (or ignorance mitigating), truth can easily emerge: we have to remember that Peirce sometimes contended that abduction "come to us as a flash. It is an act of insight" [32, 5.181] but nevertheless possesses a mysterious power of "guessing right" [32, 6.530]. Consequently abduction, preserves ignorance, in the logical sense I have illustrated above, but also can provide truth because has the power of guessing *right*. We have also contended that in the logical framework above the inference to the best explanation—if considered as a truth conferring achievement justified by empirical approval—cannot be a case of abduction, because abductive inference is instead constitutively ignorance-preserving.

If we say that truth can be reached through a "simple" abduction (not intended as involving an evaluation phase, that is coinciding with the whole inference to the best explanation, fortified by an empirical evaluation), it seems we confront a manifest incoherence. In this perspective it is contended that even simple abduction can provide truth, even if it is epistemically "inert" from the empirical perspective. Why? We can solve the incoherence by observing that we should be compelled to consider abduction as ignorance-preserving only if we consider the empirical test *the only way* of conferring truth to a hypothetical knowledge content. This clause being accepted, in the framework of the technical logical model of abduction I have introduced above the ignorance preservation appears natural and unquestionable. However, if we admit that there are ways to accept a hypothetical knowledge content different from the empirical test, for example taking advantage of special knowledge-enhancing heuristics—simple abduction is not necessarily constitutively ignorance-preserving: in the end we are dealing with a disagreement about the nature of *knowledge*, as Woods himself contends. As I have indicated at the end of the previous subsection, those who consider abduction as an inference to the best explanation—that is as a truth conferring achievement involving empirical evaluation—obviously cannot consider abductive inference as ignorance-preserving. Those who consider abduction as a mere activity of guessing are more inclined to accept its ignorance-preserving character.

However, *we are objecting that abduction—and so the possible heuristics that substantiate it—is in this last case still knowledge-enhancing*.

At this point two important consequences concerning the meaning of the word *ignorance* in this context have to be illustrated:

1 abduction, also when intended as an inference to the best explanation in the "classical" sense I have indicated above, is always *ignorance-preserving* because abduction represents a kind of reasoning that is constitutively provisional, and you can

withdraw previous abductive results (even if empirically confirmed, that is appropriately considered "best explanations"), in presence of new information. From the logical point of view this means that abduction represents a kind of nonmonotonic reasoning, and in this perspective we can even say that abduction interprets the "spirit" of modern science, where truths are never stable and absolute. Peirce also emphasized the "marvelous self-correcting property of reason" in general [32, 5.579]. So to say, abduction incarnates the human perennial search of new truths and the human Socratic awareness of a basic ignorance which can only be attenuated/mitigated. In sum, in this perspective abduction always preserves ignorance because it reminds us we can reach truths that can always be withdrawn; ignorance removal is at the same time constitutively related to ignorance regaining;

2 even if ignorance is preserved in the sense I have just indicated, which coincides with the spirit of modern science, abduction is also knowledge-enhancing because new truths can be and "are" discovered which *are not necessarily best explanations intended as hypotheses which are empirically tested*.

A similar argumentation, which resorts to better explain the conundrum of abduction as ignorance-preserving, is provided by Woods, who notes that some philosophers accept the Gabbay-Woods schema (GW-schema) for abduction but at the same time dislike its commitment to the ignorance-preservation claim. Woods' answer resorts to say that this hesitancy flows from how those philosophers *epistemologically* approach the general question of knowledge. It is not logic of abduction in question but the epistemologically adopted perspective [44, Chap. 10]. I have just said that knowledge can be attained in the absence of evidence; there are propositions about the world which turn to be true by virtue of considerations that lend them no evidential/empirical weight. They are true beliefs that are not justified on the basis of evidence. Is abduction related to the generation of knowledge contents of this kind? Yes it is.

Abduction is guessing reliable hypotheses, and humans are very good at it; abduction is akin to truth: it is especially in the case of empirical scientific cognition that abduction reveals its more representative epistemic virtues, because it provides hypotheses, models, ideas, thoughts experiments, etc., which, even if *devoid of initial* evidential support, constitute the fundamental rational building blocks for the generation of new laws and theories which only later on will be solidly empirically tested.

In the following subsections of this study I aim at illustrating this intrinsic character of abduction, which shows why we certainly can logically consider it a kind of ignorance-preserving cognition, but at the same time a cognitive—heuristic—process that can enhance knowledge at various level of human cognitive activities, even if the empirical evaluation lacks. Consequently, heuristic abductive processes can be considered knowledge-enhancing in themselves. In a previous article [26] I have shown that Peirce, to substantiate the truth-reliability of abduction—which coincides with its "ampliative" character, as illustrated in standard literature—provides philosophical and evolutionary justifications; furthermore, I have also illustrated some actual examples of knowledge-enhancing abductions active in science, that nevertheless are evidentially inert, such as in the case of guessing the so-called "conventions". They are extremely important in physics, evidentially inert fruits of abduction—at least from the point of view of their impossible falsification—but nevertheless knowledge-enhancing.

3.2 *Knowledge-Enhancing Through Models: Against Fictionalism*

I will examine in the present subsection that in science we do not have to consider the process of abducing models as a process of abducing fictions: as the reader can easily guess, this clarification will be intertwined with the aim of individuating knowledge-enhancing functions of abduction, even when clearly and immediately seen as evidentially inert.

Let us start with an example still provided by Woods, who illustrates the case of Planck's abduction as a case in which the epistemologist could see an active function of the so-called "fictions":

> When in his quest for a unified treatment of the laws of black body radiation, Planck thought up the quantum hypothesis, it was a proposition for which there wasn't a shred of antecedent evidence and none at all abduced by its presence as antecedent in the subjunctive conditional on which its provisional conjecture was based. Planck thought that the very idea of the quantum was bereft of physical meaning. It is no condition on abductive adequacy that abduced hypotheses turn out well at experimental trial. There are more things whose truth was a reasonable thing to conjecture than actually turn out to be true. [...] In some sense, the quantum hypothesis was down to Planck. Planck was the one who thought it up. Planck was the one who selected it for provisional engagement in a suitably adjusted physics. Some philosophers might see in these involvements a case for fictionalism [14].

"Planck was the one who thought it up", in this important creative event of the history of science.

Not only in the case of key hypotheses like the one proposed by Planck, but also in the case of *models* that are built as "epistemic mediators" inside a more extended process of scientific cognition, it is unlikely to admit they are abduced fictions, surely not in the minimal unequivocal sense of the word as it is adopted in the literary/narrative frameworks. Indeed, 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 and idealizations, to the more recent, fictions, credible worlds, missing systems, as make-believe, parables, as functional, as epistemic actions, as revealing capacities.[15] This proliferation of explanatory metaphors is amazing, if we consider the huge quantity of knowledge

[15] An illustration of the main problems of fictionalism and a reference to the current literature on the subject is given in my recent [25].

on scientific models that had already been produced both in epistemology and in cognitive science. Some of the authors involved in the debate on fictionalism are also especially engaged in a controversy about the legitimacy especially of speaking of fictions in the case of scientific models.

Even if evidentially inert in themselves, I think that the abduced models, both in scientific reasoning—where heuristics are often at play—and in human perception, cannot be considered as neither mere fictions, simple surrogates or make-believe, nor they are unproblematic idealizations. 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 abductive 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.[16] The abduction of these models in science is epistemically productive, models are just inert in the perspective of a *direct* empirical significance but they play a "causal" role in generating it: scientific models can be empirically false, but they are not fictions, instead they are knowledge-enhancing devices, which play an important role in reaching empirically fecund knowledge. It is clear here we are dealing with cases in which abduction is not ignorance-preserving.

Suárez [41] 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 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)[17]: the problem is that in these cases models, however, are not fictions—at

[16] In this last sense the capacity of scientific models to constitute new empirical domains and so new *empirical 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.

[17] It has to be added that Suárez does not certainly conflate scientific modeling with literary fictionalizing. He 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 [41, 35]—in the framework of an envisaged compatibility of "scientific fiction" with realism. This epistemological acknowledgment is not often present in other followers of fictionalism.

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.[18] Two consequences derive:

- 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* role of modeling in the central *abductive* creative processes, when new conceptually more or less revolutionary perspectives are advanced;
- 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 have illustrated in more detail in [25, Sect. 3.1]: abstractness and ideality cannot be solely related to empirical inadequacy and/or to theoretical incoherence [41, p. 168], in a static view of the scientific enterprise.

The considerations I have just illustrated show that model-based abduction in science is not truth preserving. Nevertheless, in the received view above, even if the guessed scientific models seem left in epistemic sufferance, scientific models cannot be considered works of fictions, and so heuristics in science acquire a special status. At this point it can be said that one thing is to abduce fictions, like in the case of creations in literature, another thing is to abduce models in empirical science. Abducing fictions in literature is also certainly knowledge-enhancing—like it is the case of scientific models—because we cannot surely imagine that literature does not provide knowledge of some kind. Moreover, how ignorance-preservation is at stake in these two cases? In the first case ignorance preservation is related to an *esthetic* failure—a fictional character can be a literary failure, discarded by the author herself—in the second one to the possible experimental discredit, which would lead to the consequent lack of *rational* success of scientific enterprise. However, no need of using in the second case the word fiction: scientific models cannot be fictions.

However, we have also to remember that—normally—abductive processes to new concepts and models in literature and in science have to also be seen as a *continuum*—a sequence—of guessed hypotheses: in both cases if the production of an intermediate abductive hypothetical model fails (and so it is abandoned), this step *could still be seen as a significant cognitive achievement* if it is useful to provide some *crucial* new information later on, useful to produce further "successful" hypotheses (for example to provide, respectively, Anna Karenina or Bohr' planetary model of the interior of atom). In this light we can say that the failed abductions were not completely esthetically or epistemically inert, because they facilitated the subsequent processes of hypothesis generation until the successful one.

---

[18] I discussed the role of chance-seeking in scientific discovery in [22].

3.3 Heuristic Abductive Steps: Dynamic Versus Static View of Scientific Models

In the perspective of the previous analysis we can further deepen in detail the process of abducing a sequence of possible hypotheses in a segment of scientific reasoning. This kind of process is often considered a *heuristic strategy*. For example, Hintikka stresses the strategic nature of abductive adoption of hypotheses inside a cognitive process:

> […] a strategic justification does not provide a warrant for any one particular step in the process. Such a particular step may not in any obvious way aid and abet the overall aim of the inquiry. For instance, such a step might provide neither any new information relevant to the aim of the inquiry nor any new confirmation for what has already been established, and yet might serve *crucially* the inquiry—for instance, by opening up the possibility of a question whose answer does so. Furthermore, notwithstanding the views of reliabilists, the idea of a nonstrategic justification that they choose is not only mistaken but in the last analysis incoherent. From the theory of strategic processes misleadingly labeled game theory, it is known that what can be valuated (assigned "utilities" to) are in principle only strategies, not particular moves. Hence a theory of epistemic processes that operates with "warrants" for particular belief changes or other things that can be said of particular moves in our "games" of inquiry is inevitably going to be unsatisfactory in the long run. One of the many things that Peirce's use of the term "hypothesis" can serve to highlight is precisely the strategic character of any justification of abduction. Being strategic, such justification does not per se lend any reliability to the outcome of some particular abductive inference. This outcome has the status of a hypothesis. Whatever reliability it may possess has to be established by the inductive component of inquiry [12, pp. 57–58].

We can agree with Hintikka that it is certainly true that we cannot have "warrants" at the level of strategic justification of particular steps in the process, and that it is also obviously true that the reliabilists are wrong suggesting the idea of a non-strategic justification. In my perspective it is the conclusion provided by Hintikka that is not satisfactory: "Being strategic, such justification does not per se lend any reliability to the outcome of some particular abductive inference. This outcome has the status of a hypothesis. Whatever reliability it may possess has to be established by the inductive component of inquiry". Hence, for example in the case of a scientific model abductively guessed, we would have to conclude—following Hintikka—that it is not reliable to the outcome of the cognitive process, indeed we have to wait for the empirical "judgment". Does this mean that every abductive guess heuristically reached is damned to be ignorance-preserving if evidentially inert? I do not think so.

To solve this problem a remark about the need of avoiding a confusion between a static and a dynamic view of scientific cognition has to be addressed. Indeed I think it is misleading to analyze the heuristic activity of abducing models in science by adopting a confounding and unclear mixture of static and dynamic aspects of the scientific enterprise. Temporal features of cognition count in understanding abduction. Scientific abduced models in a static perspective (for example when inserted in a textbook or in a text concerning history of empirical science) certainly appear—but just appear—justified "by the inductive component of inquiry", that is only in the light of the successful empirical evaluation that has been finally performed: in this case the *instrumental*—and fruitful—character of the abduced models becomes manifest, but their *constitutive* function disappears.

Please imagine that you are isolating a single moment of a dynamic creative scientific process, forgetting what will happen or happened later on, in the last case for example as it is testified in a historical narrative or in a textbook. Contrarily to the previous static view, some—the creative ones—of the abduced scientific models, once seen inside the living dynamics of scientific cognition[19] actually appear to be *explicit*, *reproducible*, and *constitutive* machineries built and manipulated to the gnoseological aim of reaching a final overall scientific result empirically evaluated, a result *not yet available*: but we have a result, the only result we have is the intermediate one, the model. The final knowledge just results not yet available because the target system and its complicated experimental apparatuses have *not yet* been built. The problem is that our snapshot of the "single moment" shows to us that these final outcomes will be built *only* thanks to the gift in terms of subsequent *knowability* provided by the intermediate models themselves. In few words: the models at play are *creative*, because they positively "establish" the root that leads to the empirical success. In a sense, their creativity "is" their reliability, they do not need further strategic or not strategic justification or warrants. If we do not acknowledge this—Kantian, I would say—aspect, we are not able to befittingly and honestly understand what abduction is as a knowledge-enhancing cognitive device.

When Hintikka complains that the abductive steps which lead to intermediate models *cannot* have "warrants" at the level of a strategic justification, and also at the level of non strategic justification, in my perspective we can relieve ourselves of this burden of epistemic sufferance just acknowledging we are dealing with *creative* models. If we only see models in empirical science in the light of the future achieved empirical success we obviously see them just as *provisional* guesses, *devoid* of justification and still and intrinsically looking for it. On the contrary, they are occasionally justified by themselves—abductively—just because creative, and so *constitutive* of a fruitful epistemic "heuristic cognitive travel". In sum, coming back to the main issue we are dealing with in this article, those models are sometimes knowledge-enhancing at their level, even if *locally* evidentially inert.

3.3.1 Fictions Versus Infinite Falsehoods

Let us reconsider in this perspective the problem of models as fictions. If we consider that the abduced models—in science—are fictions they are certainly evidentially inert; unfortunately they would be also ignorance preserving, because they will lack, as fictions, the capacity to produce a kind of intermediate knowledge endowed with the *epistemological virtue* of rationality. Woods furnishes a further

[19] Which, by the way, is the key topic of epistemology at least since Karl Popper and Thomas Kuhn.

crucial insight on the superfluity of speaking of fictions in science adopting a useful distinction between what he calls infinite (forlon) falsehoods and fictions. Given the fact that fictions detonate[20] and infinite falsehoods do not, infinite falsehoods cannot be fictions. At the same time his strict argumentation provides a further justification of the knowledge-enhancing status of scientific models. He thinks that the detonation question for infinite falsehoods (for example models that involve infinite populations in biology, that, as I have already said, many epistemologists consider fictions)

> […] is a trick question. It is a logical commonplace that, unlike truth, falsity is not preserved under consequence. How surprising can it be, then, that when $T \vdash O_i$ holds, the falsity embedded in T is not passed on to the O_i? The very fact of T's empirical adequacy precludes the detonation of its falsities. It is precisely here that fictionality's explosiveness achieves a grip. Since detonation is not a problem for falsely tinctured T_s, fictions are not required to fix it. Yet if fictions were called into play, they would create a detonation problem for T, and would guarantee that it could not be solved. For, again, detonation precludes empirical adequacy. I take this to be a serious discouragement of the fictionalist programme for science [14, p.29].

Again, if the O_i of an empirically adequate T are not derivable

> […] in the absence of $T's$ infinitely remote falsehoods, then $T's$ connection to those O_i cannot be grounding. T cannot be said to have demonstrated those consequences or to have provided a reason that supports them. This is a puzzle. But suppose, now, that fictions were called into play with a view to solving it. Then T wouldn't be empirically adequate. (Fictionality detonates.) The grounding question asks how T can be empirically adequate if it doesn't lend grounding support to the O_i in virtue of which this is so. But if fictions are let loose here, the empirical adequacy of T is lost. The grounding question wouldn't arise (*ibid*.)

Another—more epistemologically oriented—explanation of the issue at stake is given by Kuorikoski and Lehtinen [18, p. 121], who contend that: "The epistemic problem in modelling arises from the fact that models always include false assumptions, and because of this, even though the derivation within the model is usually deductively valid, we do not know whether our model-based inferences reliably lead to true conclusions". The problem is that false premises (also due to the presence in models of both substantive and auxiliary assumptions) are not exploited in the cognitive process, because, in various heuristic steps, only the *co-exact* ones are exploited. For example, the notion of co-exact proprieties, introduced by Manders [27], is worth to be further studied in fields that go beyond the realm of discovery processes of classical geometry, in which it has been nicely underscored. Mumma [29, p. 264] illustrates that Euclid's diagrams contribute to proofs only through their co-exact properties. Indeed "Euclid never infers an exact property from a diagram unless it follows directly from a co-exact property. Exact relations between magnitudes which are not exhibited as a containment are either assumed from the outset or are proved via a chain of inferences in the text. It is not

[20] "The detonation question: How widely spread in a theory T is the alethic impact of its ineliminable idealizations? How contagious is the property of infinite falsehood?" [14, p. 19]

difficult to hypothesize why Euclid would have restricted himself in such a way. Any proof, diagrammatic or otherwise, ought to be reproducible".

Moreover, as I have already noted, some false—eventually abduced—assumptions are considered as such *only if* seen in the light of the still "to be known" target system, and so they appear false only in a post hoc analysis, but they are perfectly true—and so knowledge-enhancing—in the model itself in its relative autonomy during the smart *heuristic* cognitive process related to its exploitation. Falsities in one perspective can be considered (local) truths in another one. So various aspects of the model are the legitimately true basis for the subsequent exploration of its behavior and performance of further abductions or other inferences to statements concerning the target system. I agree with Morrison: "I see this not as a logical problem of deriving true conclusions from false premises but rather an epistemic one that deals with the way false representations transmit information about concrete cases" [36, p. 111].[21]

Morrison [36] is certainly not inclined to see models as fictions because she emphasizes that in science they are specifically related to ("finer graded") ways of understanding and explaining "real systems", far beyond their more collateral predictive capabilities and their virtues in approximating. She indeed further clarifies that the models which is appropriate to label as *abstract* resist—in the so-called process of de-idealization—corrections or relaxing of the unrealistic assumptions (such as in the case of mathematical abstractions or when models furnish the sudden chance for the applicability of equations), because they are "necessary" to arrive to certain results: models are not redundantly required for the derivation of the predictions by which gain a contact with the observational level. The fact that in these models "relevant features" are subtracted to focus on a single—and so isolated—set of properties or laws, as stressed by Cartwright [4], is not their central quality, because what is at stake is their capacity to furnish an overall new depiction of an empirical (and/or theoretical, like in case of mathematics or logic) framework: "[…] We have a description of a physically unrealizable situation that is required to explain a physically realizable one" (p. 130). In a similar vein Woods nicely concludes "A central role for empirically forlorn representations in model-based science is the establishment of non-probative premiss-conclusion linkages in ways that set up their conclusions for empirical negotiation at the checkout counter" [43]. The cognitive situations just described still reflect those cases of abductive successes in model-based reasoning, which are simply not probatively successful from the ultimate empirical point of view, but, as we are trying to demonstrate, not necessarily devoid of a knowledge-enhancing status.

[21] Further information about the problem of the mapping between models and target systems through *interpretation* are provided by Contessa [5, p. 65]—interpretation is seen as more fundamental than surrogative-reasoning: "The model can be used as a generator of hypotheses about the system, hypotheses whose truth or falsity needs to be empirically investigated". By using the concept of interpretation (analytically and not hermeneutically defined) the author in my opinion also quickly adumbrates the creative aspects in science, that coincide with the fundamental problem of model-based and manipulative abduction.

Moreover, many models, easier to define, which is better to classify as *idealizations*, allow "[…] for the addition of correction factors that bring the model system closer (in representational terms) to the physical system being modelled or described" [36, p. 111]. It is for example the case of simple pendulum, where we know how to add corrections to deal with concrete phenomena. Idealizations distort or omit properties, instead abstractions introduce a specific kind of representation "that is not amenable to correction and is necessary for explation/prediction of the target system" (p. 112), and which provides information and transfer of knowledge. Morrison's characterization of scientific models as abstract is in tune with my emphasis on models as creative and so accepted as *constitutive*, beyond the mere role played by models as idealizations, which instead allow corrections and refinements.[22] In this perspective, "abstract" models, either related to prepare and favor mathematization or directly involving mathematical tools, have to be intended as poietic ways of producing new intelligibilty of the essential features of the target systems phenomena, and not mere expedients for facilitating calculations. If idealization *resembles* the phenomena to be better understood, abstract models abductively *constitute* the resemblance itself.

It is not that "fictions provide inferential shortcuts in models; and the fact that this is the main or only reason for their use distinguishes them as fictional" [35, p. 239], even if Vaihinger would agree with this functionalist perspective on fictions.[23] Indeed, even if it is not decisive to say "that the inferential characterisation provides a way to distinguish precisely scientific from non-scientific uses of fiction", the abduced models used in non-scientific practices may also trigger inferences, and the problem here is more fundamental. In science, models are not used and intended as fictions, they are just labeled as fictions because of a juxtaposition of some recent philosophers of science, who certainly in this way render the scientific enterprise more similar to other more common modes of human cognition: after all fictions are ubiquitous in human cognition, and science is a cognitive activity like others. Unfortunately science never aimed to abduce "fictions" at the basic levels of its activities, so that the recent fictionalism 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 heuristic finalities of science, shading in a kind of debate about entities and their classification that could remind of medieval scholasticism. Abduction is a widespread cognitive activity, like its logical models teach to us, and so it is the activity of abducing models, but this huge extension paradoxically furnishes a further reason for the philosopher, the epistemologist, and the cognitive scientist to study,

[22] On the constitutive vs. descriptive role of models cf. also [39].

[23] I have to add that Suárez does not defend the view according to which models are fictions: even if he defends the view that models contain or lead to fictional assumptions, he rejects the identification of models and fictions, preferring instead to stay "quietist" about the ontology of models, and focusing rather on modeling as an activity—see in particular [40].

differentiate, and respect the various types of knowledge, beliefs, and levels of truth and/or rationality more or less involved.

3.3.2 Heuristics in Deductive Reasoning

Finally, we have to acknowledge that model-based abductive cognitive constituents are also present in deductive reasoning, and so in mathematics and logic. Peirce himself was clearly aware, speaking of the model-based aspects of 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" [32, 3.363].

Also Hintikka clearly shows the "embarrassing" presence of fruitful abductive creative moments in deduction, which are invaded by strategic/heuristic hypothetical interventions crucial to proceed and reach the final results, and that of course are evidentially inert. Also in deduction, the presence of abductive heuristic events coincides with their knowledge-enhancing character: here too these *strategic* aspects reflect the pure—productive—conjectural element of abductive inference and its capacity to guessing right.[24] Hintikka clearly points out the abductive nature of the inferential phase in which the existential quantifier is introduced. This case is in turn related to his emphasis on the *strategically* positing of the "right questions" which "depend on one's ability to anticipate their answers" [12, p. 55]:

> […] the very same sentence can serve as the presupposition of a question and as the premise of a deductive step. For instance, an existential sentence of the form
>
> (1) $(\exists x)S[x]$
>
> can serve either as the presupposition of the question
>
> (2) What (who, when, where,…), say x, is such that $S[x]$?
>
> or as the premise of an existential instantiation that introduces a John Doe—like "dummy name" of an "arbitrary name", say β. In the former case, the output of the relevant step is a sentence of the form
>
> (3) $S[b]$
>
> where b is a singular term—for instance, a proper name. In the latter case, the output is of the form
>
> (4) $S[\beta]$

[24] These strategic moves correspond to particular forms of abductive reasoning. In Beth's method of semantic tableaux the strategic "ability" to construct impossible configurations is undeniable [13, 30]. Also Aliseda [2, 1] provides interesting use of the semantic tableaux as a constructive representation of theories, where abductive expansions and revisions, derived from the belief revision framework, operate over them. The tableaux are so viewed as a kind of reasoning (non-deductive) where the effect of "deduction" is performed by means of abductive strategies.

> Here, (4) differs from (3) only by having a dummy name, whereas in (3) there was a real name.
>
> […] It seems to me that Peirce had an intuitive understanding of this type of similarity between abductive and deductive inferences. […] These similarities between questions (abductive steps) and logical inferences (deductive steps) are purely formal, however. An epistemological assimilation of the two to each other on the mere basis of such formal similarities would be irresponsible. The crucial insight here is that behind these formal similarities there lies a remarkable strategic similarity [12, pp. 53–54].

The strategic similarity resorts to the need of the reasoner of using one of the propositions that are available to her as presuppositions or as premises, *both* in the case of abduction and of deduction: "Which sentence or sentences should I use as the premise or as the premises of a deductive inference? It can be shown that the most sensitive strategic question in deduction is: Which sentence should I use first as the premise of an existential instantiation or its generalization, functional instantiation? […] If the inquirer is reasoning empirically (interrogatively), the next strategic question is: Which one of the available sentences should I use as the presupposition of a why question? These candidate sentences are the very same ones that could be used as premises of existential instantiations, suitably generalized. Neither question admits in general of a mechanical answer, in the sense that there is in neither case any recursive function that always specifies an optimal choice. […] In this sense, the strategic principles of abductive reasoning, interpreted as I have done, *are the same as the strategic principles governing deduction*" ([12, p.54], emphases added).

Analogously, Kant himself returned to the tradition which envisions processes of analysis and synthesis operating in geometrical reasoning.[25] In proving geometrical theorems, it suffices to consider the individual given figure, i.e., the figure of which the antecedent of the proposition speaks. Then, it is necessary to provide a "preparation" or a "construction"—or "machinery"—in Greek: (κατασκευή) in order to be able to conduct the proof (that is to complement the given figure by drawing new lines, circles, and other diagrams) [10, p. 202]. Moreover, Hintikka recalls, Socrates himself traced figures on the basis of *Meno*'s reasoning and these figures are the same starting points of the slave's analysis:[26] "To some extent, Kant also seems to have thought of another part of the proposition as being synthetic, namely, the setting-out or *ecthesis* (ἔχθεσις) which immediately follows the general enunciation of the proposition in question and in which the geometrical entities with which the general enunciation deals are 'set out' or 'exposed' in the form of a particular figure" (*ibid.*, p. 209). Indeed Kant considers the true demonstration (ἀπόδειξις) to be analytic "[…] which follows the auxiliary construction and in which no new geometrical objects are introduced. In this ἀπόδειξις we merely analyse, in a fairly literal sense of the word, the figure introduced in the ecthesis and completed in the 'construction' or 'machinery' " [10, p. 209]. This definition of construction means that Kant saw the central aspect of mathematics in the

[25] On the relationship between analysis and synthesis in the history of geometry and in particular in the history of philosophy of geometry refer to the many remarks given in [19, 31]. Cf. also [6].

[26] On this Plato's dialogue cf. [21, Chapter 6, Sect. 1].

introduction of representatives of individuals (intuitions, that is the free individual terms emphasized by Hintikka) which instantiate general concepts. The rule of existential instantiation offers a very typical case. The presence in this process of those manipulative aspects of cognition I will stress below is evident.Taking advantage of the previous notes about the abductive aspect of mathematical reasoning (and discovery) and logical deduction let us come back to the problem of abduction and of its presumptive ignorance-preserving character. I have many times stressed in my works that manipulative abduction (see above, footnote 14), which is widespread in scientific reasoning, is a process in which a hypothesis is formed resorting to a basically extra-theoretical and extra-sentential behavior that aims at creating communicable accounts of new experiences to the final aim of integrating the successful results 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 for example in a neural engineering laboratory, useful to make observations and "experiments" to transform one cognitive state into another to discover new properties of the target systems. Manipulative abduction refers to those more unplanned and unconscious action-based cognitive processes I have characterized as forms of "thinking through doing".[27] It is clear that manipulative abduction in science basically deals with the handling of external models in their intertwining with the internal ones. Consequently, even if related to experiments occasionally performed with the help of external models sometimes mediated by artifacts, manipulative abduction has to be considered—obviously in mathematics but also in the case of empirical science, evidentially inert, even if of course not necessarily ignorance-preserving, as I have tried to demonstrate in this paper.

I have contended that manipulative abduction is also active in mathematics. For example, we have already seen that Peirce, in the case of mathematics, speaking of the model-based aspects of this kind of deductive reasoning, hypothesized there is an "experimenting upon this image [the external model/diagram] in the imagination", so showing how human geometrical imagination is always triggered by a kind of prosthesis, the external model as an "external imagination". Analogously, taking advantage of a fictional view on models and of the pretence theory Frigg [8, p. 266 ff.] interestingly sees imagination as an authorized intersubjective heuristic game of make-believe sanctioned by the "prop" (an object, for example material models, movies, paintings, plays, etc.) and its rules of generation. This theory also works as a metaphor of abductive processes, in terms of some concepts taken from the theory of literary and artistic fictions: again, I think that it is neither necessary to adopt a fictionalist view in the case of science, nor the pretence theory

[27] I have to note that manipulative abduction also happens when we are *thinking through doing* (and not only, in a pragmatic sense, about doing). This kind of action-based cognition can hardly be intended as completely intentional and conscious.

adds something relevant to the issue and, moreover, this kind fictionalism would obscure the knowledge-enhancing role of abduction we are describing in this paper.

Analogously, in the example I have illustrated in [25], concerning the exploitation of concrete/external models (for example in vitro or computational) in a scientific lab, scientists do not pretend anything and are not engaged in the relative make-believe process, if not in the trivial sense that almost every human intersubjective interplay can be seen as such. The in vitro networks of cultured neurons of that case or the Peircean Euclidean diagram used by the ancient Greek geometers are just the opposite of a mere fiction or of a generic make-believe interplay, they are instead more or less mimetic (possibly creative and so enhancers of new knowledge not already available) external models—reached through manipulative abduction—which are expected to heuristically provide reliable information about the target system. They aim at abductively discovering some new representations about the neurons in the first case and about the pure concepts of geometry in the second.

As I have anticipated, here we see that, even in the mathematical discovery processes, manipulative abduction is based on the heuristic interplay between internal and external representations (not only diagrams, but also written proofs, etc.): the final result is an abductive hypothesis which assumes the clothes of a Kantian "stipulation", endowed with epistemic virtues, that same "productive" stipulation, squarely evidentially inert, we have seen at work in the case of heuristic steps of model-based science and in deductive reasoning.

Finally, it is important to explicitly emphasize the intrinsic relativity of the status of concepts like truth, rationality, knowledge, ignorance: their reciprocal entanglement tends to reciprocally depict the respective meanings. Here an example: it has to be said that successful abductions that are performed at the moral level I have mentioned above immediately acquire a deontological status. They are *epistemically inert* even if they increase something that we can certainly call "knowledge": the "moral" knowledge human individuals need in a given situation. The use of the word knowledge depicts the meaning of the word ignorance: at least under the perspective of the last moral case, the abductions involved *are not* ignorance-preserving, because do not preserve the subjective moral ignorance in front of the problems of what I have called "military intelligence"[28] at stake. Nevertheless, the abduced hypothetical knowledge in the case of these—primarily moral—endeavors can be easily seen a piece of "false" knowledge, from the empirical and/or rational point of view, but still active and efficient, and in this case we are legitimated to call the involved abductions as basically ignorance-preserving. Indeed, common moral knowledge of beings like us is not intrinsically truth-sensitive, at least in the sense that the word truth acquires in rational and scientific settings.

[28] I have illustrated the role of abduction in "military intelligence" in [24, Chap. 2], where I have extendedly treated the relationship between cognition and violence.

4 Conclusion

The status of abduction is very controversial. When dealing with abductive reasoning misinterpretations and equivocations are common. What did Peirce mean when he considered abduction both a kind of inference and a kind of instinct or when he considered perception a kind of abduction? Does abduction involve only the generation of hypotheses or their evaluation too? Are the criteria for the best explanation in abductive reasoning epistemic, or pragmatic, or both? Does abduction preserve ignorance or extend truth or both? What did Peirce mean when—anticipating the heuristic perspective—considered an abductive hypothesis as the fruit both of "the different elements [that] were in our minds before" and of "the idea of putting together what we had never before dreamed of putting together"?

The paper has tried to answer these questions centering the attention to the so-called ignorance-preservation character of abduction, contrasted with its knowledge-enhancing capacity, such as it is expressed by its heuristic features. I have contended that even if, certainly, abductive reasoning can be considered a response to an ignorance-problem, nevertheless, through abduction, knowledge can be enhanced, even when abduction is not considered an inference to the best explanation in the classical sense of the expression, that is an inference necessarily characterized by an empirical evaluation phase.

To study this theoretical conundrum I exploited my *eco-cognitive model* (EC-model) of abduction and illustrated that Hintikka is wrong when he contends that the abductive steps which lead to intermediate models *cannot* have "warrants" at the level of a heuristic justification: I contended that we can relieve ourselves of this burden of epistemic sufferance just acknowledging that in various heuristic cases we are dealing with *creative* models. If we only see the heuristic use of models in empirical science in the light of the future achieved empirical success we obviously see them just as *provisional* guesses, *devoid* of justification and still and intrinsically looking for it. On the contrary, they are occasionally justified by themselves—abductively—just because creative, and so *constitutive* of a fruitful epistemic "heuristic cognitive travel". I have also consequently shown that if we see heuristics, in the case of science, in a static perspective, they appear unwarranted to the epistemologist, but this presumptive groundless character vaporizes if a dynamic perspective is assumed. Abductive heuristics in scientific model-based reasoning do not constitute a questionable process of guessing fictions. Finally, I have illustrated that, also in deduction, the presence of abductive heuristic events coincides with their knowledge-enhancing character: here too these strategic aspects reflect the pure—productive—conjectural element of abductive inference and its capacity to guessing right.

References

1. Aliseda, A.: Abductive Reasoning Logical Investigations into Discovery and Explanation. Springer, Berlin (2006)
2. Aliseda, A.: Seeking explanations: abduction in logic, philosophy of science and artificial intelligence. PhD thesis, Institute for Logic, Language and Computation, Amsterdam (1997)
3. Angius, N.: Towards model-based abductive reasoning in automated software testing. Logic J. IGPL (2013)
4. Cartwright, N.: Nature's Capacities and Their Measurement. Oxford University Press, Oxford (1989)
5. Contessa, G.: Scientific representation, interpretation, and surrogative reasoning. Philos. Sci. **74**, 48–68 (2007)
6. De Angelis, E.: Il metodo geometrico nella filosofi a del Seicento. Istituto di Filosofia, Università degli Studi di Pisa, Pisa (1964)
7. Feyerabend, P.: Against Method. Verso, London (1975)
8. Frigg, R.: Models and fictions. Synthese **172**, 251–268 (2010)
9. Gabbay, D. M., Woods, J.: The reach of abduction. North-Holland, Amsterdam (2005) (Volume 2 of A Practical Logic of Cognitive Systems)
10. Hintikka, J.: Logic, Language-Games and Information. Clarendon Press, London (1973)
11. Hintikka, J.: The place of C.S. Peirce in the history of logical theory. In: Brunning, J., Forster, P. (eds.) The Rule of Reason. The Philosophy of Charles Sanders Peirce. University of Toronto Press, Toronto (1997)
12. Hintikka, J.: Socratic Epistemology. Explorations of Knowledge-Seeking by Questioning. Cambridge University Press, Cambridge (2007)
13. Hintikka, J.: What is abduction? The fundamental problem of contemporary epistemology. Trans. Charles S. Peirce Soc. **34**, 503–533 (1998)
14. Woods, J.: Against fictionalism. In: Magnani, L. (ed.) Model-Based Reasoning in Science and Technology. Theoretical and Cognitive Issues, pp. 9–49. Springer, Heidelberg (2013)
15. Kakas, A., Kowalski, R.A., Toni, F.: Abductive logic programming. J. Logic Comput. **2**(6), 719–770 (1993)
16. Kowalski, R.A.: Logic for Problem Solving. Elsevier, New York (1979)
17. Kuipers, T.A.F.: Abduction aiming at empirical progress of even truth approximation leading to a challenge for computational modelling. Found. Sci. **4**, 307–323 (1999)
18. Kuorikoski, J., Lehtinen, A.: Incredible worlds, credible results. Erkenntnis **70**, 119–131 (2009)
19. Lakatos, I.: Proofs and Refutations. The Logic of Mathematical Discovery. Cambridge University Press, Cambridge (1976)
20. Mackonis, A.: Inference to the best explanation, coherence and other explanatory virtues. Synthese **190**, 975–995 (2013)
21. Magnani, L.: Abduction, Reason, and Science. Processes of Discovery and Explanation. Kluwer Academic/Plenum Publishers, New York (2001)
22. Magnani, L.: Abduction and chance discovery in science. Int. J. Knowl. Based Intell. Eng. **11**, 273–279 (2007)
23. Magnani, L.: Abductive Cognition. The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning. Springer, Heidelberg (2009)
24. Magnani, L.: The Interwining of Morality, Religion, and Violence: A Philosophical Stance. Springer, Heidelberg (2011)
25. 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)

26. Magnani, L.: Is abduction ignorance-preserving? Conventions, models, and fictions in science. Logic J. IGPL **21**(6), 882–914 (2013)
27. Manders, K.: The Euclidean diagram. In: Mancosu, P. (ed.) Philosophy of Mathematical Practice, pp. 112–183. Clarendon Press, Oxford (2008)
28. Meheus, J., Verhoeven, L., Van Dyck, M., Provijn, D.: Ampliative adaptive logics and the foundation of logic-based approaches to abduction. In: Magnani, L., Nersessian, N.J., Pizzi, C. (eds.) Logical and Computational Aspects of Model-Based Reasoning, pp. 39–71. Kluwer Academic Publishers, Dordrecht (2002)
29. Mumma, J.: Proofs, pictures, and Euclid. Synthese **175**, 255–287 (2010)
30. Niiniluoto, I.: Abduction and geometrical analysis. Notes on Charles S. Peirce and Edgar Allan Poe. In: Magnani, L., Nersessian, N.J., Thagard, P. (eds.) Model Based Reasoning in Scientific Discovery, pp. 239–254. Plenum Publishers/Kluwer Academic, New York (1999)
31. Otte, M., Panza, M.: Analysis and Synthesis in Mathematics. Kluwer Academic Publisher, Dordrecht (1999)
32. Peirce, C.S.: Collected Papers of Charles Sanders Peirce. Harvard University Press, Cambridge, MA, 1931–1958. In: Hartshorne, C., Weiss, P. (eds.) vols. 1–6; Burks, A.W. (ed.) vols. 7–8 (1931–1958)
33. Peirce, C.S.: Historical Perspectives on Peirce's Logic of Science: a History of Science. In: Eisele, C. (ed.) vols. I–II. Mouton, Berlin (1987)
34. Peirce, C.S.: Visual cognition and cognitive modeling. In Philosophical Writings of Peirce, pp. 302–305. In: Buchler, J. (ed.) Dover, New York (1955)
35. Suárez, M.: Fictions, inference, and realism. In: Woods, J. (ed.) Fictions and Models: New Essays, pp. 225–245. Philosophia Verlag, Munich (2010)
36. Morrison, M.: Fictions, representations, and reality. In: Suárez, M. (ed.) Fictions in Science: Philosophical Essays on Modeling and Idealization, pp. 110–135. Routledge, London (2009)
37. Rivera, F.D., Rossi, J.: Becker. Abduction-induction (generalization) processes of elementary majors on figural patterns in algebra. J. Math. Behav. **26**, 140–155 (2007)
38. Shanahan, M.: Perception as abduction: turning sensory data into meaningful representation. Cognitive Sci. **29**, 103–134 (2005)
39. Stefanov, A.: Theoretical models as representations. J. Gen. Philos. Sci. **43**, 67–76 (2012)
40. Suárez M. (ed.): Fictions in Science: Philosophical Essays on Modeling and Idealization. Routledge, London (2009)
41. 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)
42. Woods, J.: Ignorance, inference and proof: abductive logic meets the criminal law. In: Tuzet, G., Canale, D. (eds.) The Rules of Inference: Inferentialism in Law and Philosophy, pp. 151–185. Egea, Heidelberg (2009)
43. Woods, J.: Mathematizing knowledge (2012) (Unpublished Paper)
44. Woods, J.: Errors of Reasoning. Naturalizing the Logic of Inference. College Publications, London (2013)
45. Woods, J.: Recent developments in abductive logic. Stud. Hist. Philos. Sci. **42**(1), 240–244 (2011) (Essay Review of L. Magnani, Abductive Cognition. The Epistemologic and Eco-Cognitive Dimensions of Hypothetical Reasoning, Springer, Heidelberg/Berlin, 2009)

Heuristic Appraisal at the Frontier of Research

Thomas Nickles

Abstract How can we speed up both basic and translational scientific research without major new financial investment? One way is to speed up the process by which good proposals are funded. Another is to do a better job of identifying research that is potentially transformative. There are internal institutional barriers as well as sluggish and conservative policies in place in many government funding agencies, universities, and private firms, policies that are risk-averse and characterized by short-term accounting. While perhaps calling for transformational research, their selection procedures promote normal basic research and translational research instead. This chapter proposes that progress can be made by giving increased weight to heuristic appraisal—appraisal of the future promise of proposed research—and correspondingly less weight to confirmational appraisal—the logical and probabilistic relations between theories and data sets already on the table. Emphasis on the latter, as studied by traditional confirmation theory, is a legacy of logical positivism. Adapting a form of scenario planning from the business community is one positive suggestion.

Keywords Frontier research · Problem choice · Heuristic appraisal · Confirmation theory · Science policy · Transformational research · Innovation

1 Introduction

How can we overcome the conservative bias that increasingly discourages transformative research? In this contribution I contend that giving somewhat more weight in scientific decision-making to what I call *heuristic appraisal* (HA)—and correspondingly less weight to traditional *confirmational appraisal* (CA)—can help. Consider

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

E. Ippoliti (ed.), *Heuristic Reasoning*, Studies in Applied Philosophy, Epistemology and Rational Ethics 16, DOI 10.1007/978-3-319-09159-4_4

the *frontier problem problem*, the problem of choosing a research problem or project to work on at the research frontier—or of choosing which research proposals to fund. Truly creative scientists at the frontier face decision problems under risk and uncertainty, as do panelists and research support administrators. Unfortunately, competition for limited funds often interposes an uncreative tension between these two (or more) sets of decisions, the decisions the scientists make (or would like to make) and the decisions of the granting agency and institutional administrators.

Some scientists would like to take the kinds of risks that can lead to breakthroughs, but government funding agencies in today's industrialized democracies, while perhaps calling for transformative proposals, are risk-intolerant under political pressure neither to waste tax dollars nor to spend them on politically "edgy" projects. Corporate laboratories can be somewhat similarly constrained, especially in large firms with rigid, hierarchical, "command and control" administrative structures [47, 49]. In both cases there is pressure to produce "translational" results in the short run. I shall focus on grant support for university-based research. One frequently hears that granting organizations have become increasingly risk-intolerant, that they want to support only proposals for which useful results are practically assured. To many investigators, this means that they must already have done a good bit of the research and obtained promising results before submitting a significant request for support.

Grant proposals to government agencies are typically vetted rigorously by a centralized system of reviewers, panelists, and division directors. The process is so inefficient that it often takes as long to get a proposal through the system as the length of the proposed grant itself. Moreover, the evaluation process is often biased toward applying the same criteria to proposed work as is applied to finished work. This is typically an epistemological "realist" bias (relativized to current orthodoxy) in which it is supposed that the established theories and models of a successful, mature science are already close to the truth. Hence its failure to square with current orthodoxy often results in a "knock-off" argument against a proposal. To be sure, most such proposals probably do deserve to fail, but there are surely cases in which a more flexible approach to grant support would pay off in the long run.

Traditional philosophical accounts of scientific work do not provide helpful resources to overcome this conservatism. With few exceptions philosophers of science have treated frontier decision-making as derivative from whatever they take to be standard confirmation theory or theory of justification, with the conditions somewhat relaxed at the proposal and early research stages as compared with eventual acceptance or rejection of a "finished" product. Unfortunately, however, there is no agreement on what confirmation theory is supposed to be. Besides, judging them by criteria anchored to current orthodoxy would seem to be a poor way to evaluate potentially transformative projects.

Another way of expressing my claim is to say that granting agencies are increasingly adopting what may be called the philosopher's point of view rather than the creative scientist's point of view. Creative scientists are future-oriented in thinking about the most promising problem they can tackle next.

The traditional philosopher's point of view is retrospective, focused on the degree of support (probability, degree of truth-likeness, etc.) that a given data set confers on a theory that is already well articulated. This view was inherited from the logical positivists (of the Vienna Circle) and the logical empiricists (of the Berlin Circle), who tended to be skeptical of the epistemological relevance of non-confirmational frontier research, which they labeled "context of discovery" as opposed to "context of justification." A retrospective view is also embraced by many of today's strong, epistemological realists. They no longer share the positivists' worries about the meaningfulness of theoretical language, but they regard highly successful confirmational track records as the mark of a mature science that is now close to the truth (e.g., [4, 57]). For them, mature science is a permanent sort of normal science, so we should expect no major future surprises.

Such a perspective inadvertently fosters the layperson's view (and most legislators are laypersons) that science is a body of established truths and the technological gadgets that they make possible rather than an ongoing process of inquiry that is likely to alter present understandings significantly in the future.

One trouble with the retrospective view is that it is not retrospective enough. It goes only as far as to examine the logical or probabilistic relations among the aforementioned items already on the table and fails to consider the long-term dynamics of the history of science, and thus fails to project those dynamics onto the future. A good question for both investigators and grants providers to keep in mind is: "How likely is it that our present understandings in a given mature scientific domain will be retained by scientists 50 or 100 years into the future?" When we look at past "predictions" of prominent people about what life will be like a 100 years into the future, most of them are laughable in retrospect. Why should our guesses about the future of science to be so different?

The practical question for policy recommendations is to what degree we can turn this historical hindsight into foresight, given that forecasts of the future become rapidly unreliable as we move out from the present. Nearly all of the future will surely be beyond our current horizon of imagination. "The future" is a long time! Part of the answer, I believe, is furnished by what Mary Hesse termed "*a principle of no-privilege*, according to which our own scientific theories are held to be as much subject to radical conceptual change as past theories are seen to be" ([24, p. 264], her emphasis).

Another part of the answer is furnished by Thomas Kuhn's famous opening sentence in *The Structure of Scientific Revolutions*:

> History, if viewed as a repository for more than anecdote or chronology, could produce a decisive transformation in the image of science by which we are now possessed [29, p. 1].

Although often repeated, few writers take this (or Hesse's) statement seriously enough. These analysts seem to suppose that "history" is what is in our past and, so, ends with us—and that we are historically up to date, compared to our predecessors, by beginning to take stock of it. Whereas, barring worldwide disaster, most of the history of science, by far, lies in our future. As a historian, Kuhn himself was surely referring to what we can learn from the past. I am giving his statement the more radical meaning, as a statement about what we might call *total history* as opposed to *our* history, in line with Kuhn's own willingness to project past revolutionary dynamics onto the future. The ambiguity of 'history' (our history versus total history) lulls us to sleep. We should test our methodological and policy accounts not only against the known past but also against possible futures. Suppose we extend the historical timeline 40,000 years into the future. From this perspective we realize that our time-series of scientific developments since, say, 1600, provides a miniscule and early-biased sample of total scientific history. To sum up, we should thus take Kuhn's opening claim as equally being about our future. (I do not, however, commit myself to his overly sharp distinction between normal and extraordinary science, nor to his position on incommensurability.)

A third part of the answer attempts to remedy the just-mentioned fault by comparing scientific work to technological innovation, to the construction of technological designs, designs that face competition in the marketplace and the occasional "waves of creative destruction" [63] that we find in the world of commercial technological design and modern economic life. Again, no one expects the technology of 50 or a 100 years into the future to be much like ours today, so why should science be so different? After all, science will be driven, in part, by these technological changes. Any science that is not dynamic in something like this manner is relatively stagnant. The more creative researchers will leave for greener pastures.

Whatever the decision procedure may be between two finished hypotheses or theories and a data set, the frontier situation is different, and all research grant applications worth their salt deal with frontier research. In making research decisions at the frontier, I claim that creative scientists give substantial weight to heuristic appraisal, sometimes to the extent that HA dominates CA, which is typically thin on the ground at that point in any case. Moreover, truly creative researchers will often follow fruitful leads that challenge accepted "knowledge." Fertility can even trump expected truth when the supposedly true claim does not leave the scientists anything very interesting to do. That is why creative researchers may leave a domain they consider basically finished in order to strike out for new territory [43, 45].

The goals of ongoing research are (or should be) set by fertility-seeking more than by direct truth-seeking. This is not to deny that many practicing scientists are trying to establish the truth about something. Nor is it to say that "off the wall" ideas should be encouraged, for any violation of previously established results needs to be offset by great heuristic promise. I am not advocating irresponsibly "wild" behaviors. But it is good to remember that even in basic research supposed truths are often more valuable for the heuristic platform they provide than because they are true. After all, many of the big conceptual breakthroughs in the history of science have been seriously false, even internally inconsistent initially, or inconsistent with the received "knowledge" of the day. It remains an open question whether truth as an explicit constraint on theories or models really adds anything to the scientists' methodological toolkit in any given field, but there is no question that fertility is crucial. The dispute about truth is not one that I can engage here, but it is worth noting that models are at least as pervasive in research as are theories of any size, and most models are known from the start to be false, strictly speaking, given that they involve some form of idealization or approximation.

2 The Epistemological Importance of Innovation—and Rate of Innovation

A concern with research policy engages some of the deepest epistemological issues. It is innovation, especially transformative innovation, that drives vital enterprises of all kinds. This implies that successful enterprises are highly dynamic, self-transforming, and constantly moving into new and uncertain terrain. Many of those who discuss "conceptual change" implicitly assume a linear, teleological progress toward the truth. For them, a conceptual change is a step closer to a final representation of the universe. Once a science passes the maturity threshold, the expectation is that significant changes will become asymptotically smaller as we near the goal.[1] Yet most of the same methodologists will readily agree that today's scientific technology (detection instruments, information processing devices, etc.) will soon be obsolete as new technology is invented. Nicholas Rescher (who adopts a more dynamic conception of the future) puts the point well:

> The very structure of scientific inquiry, like an arms race, forces us into constant technological escalation where the frontier equipment of today's research becomes the museum piece of tomorrow under the relentless grip of technical obsolescence [59, p. 93].
>
> (1) We can learn about nature only by interactively pushing up against it, and what it is that it yields to us will depend on how hard we push. (2) Our view of the world accordingly changes in the course of technology-induced scientific progress. (3) The picture of nature we obtain at one level of observational and experimental sophistication becomes destabilized at the next level. (4) With scientific progress we constantly have to reconstruct the concept mechanism we use in science [60, p. 51].

I agree. New technologies, new discoveries, tame the frontiers of our predecessors, but they open up entirely new frontiers. In that respect, they present us with a significantly different universe than that of our predecessors.

I regard the sciences as extremely complex, socially embedded systems that are better viewed as a kind of Hegelian hodgepodge over which we have little lookahead and little control over future directions. Heavily centralized, rational planning is no more desirable or even possible here than in managing a national economy. Future scientific and technological innovation of a transformative sort cannot be intelligently designed—at least not directly. Just as economists do not yet fully understand the origins of wealth (see [2, 73]), philosophers and (other) science studies experts do not fully understand the origins, the sources, of knowledge.

---

[1] This is so not only for analysts who would define truth epistemically as what we get in the limit of successful inquiry but also for those strong realists who reject the epistemological reduction of truth.

The central questions that I wish to ask are these: *How can we increase the pace of innovative scientific research, especially transformative research, short of a major infusion of new resources*? *What methodological or policy steps can we take*? My general answer will be: *By increasing the opportunity for transformative fertility judgments at frontiers and for increasing the weight of those judgments that do occur*.

This answer naturally generates two subproblems: (1) *How do scientists and those involved in their support networks make fertility decisions at frontiers*? (2) *Which factors stimulate or hinder such frontier opportunities*?

My label or placeholder for the answer to the first subproblem is *heuristic appraisal* (HA). This is the evaluation of the future promise, the expected or possible fertility of anything that contributes to advancing innovative research.[2] The choices are not limited to theories. HA applies to problem choices, choice of tools, personnel choices for research teams, models, experimental design, the way in which research institutions are designed, funding choices, even hiring designs and organization of work spaces.

I shall contrast heuristic appraisal with *confirmational appraisal* (*CA*) in the sense of evaluation of the empirical and problem-solving track record that has been the mainstay of traditional confirmation theories.[3] There are many differences. Here are two of the most important. (1) HA is prospective, directed toward possible future developments, future opportunities, whereas CA is retrospective, based on past performance.[4] (2) There is thus an important modal difference—that between (perceived or judged) potentiality and actuality. I am using 'confirmation' in a sense broad enough to include success in solving conceptual as well as empirical problems, and to include explanatory power and coherence as well as predictive power.

My answer to the question, What methodological and policy steps can we take?, will result in an irony, namely, that some of the most epistemologically-relevant steps are things that demote the epistemological factors that have most interested philosophers, while promoting factors that philosophers have routinely considered "external" to epistemology.

---

[2] See [43, 45, 48, 49]. In [45] I spoke too much of *expected* fertility and gave insufficient emphasis to the *mere possibility* of breakthroughs that can drive research in crisis situations. It should also be clear that 'future promise' is intended to cover a range of possibilities, from strong to weak. Generally, 'future promise' cannot be modeled by a probability calculus. Carlo Cellucci and Emiliano Ippoliti are on a better track to speak of "plausibility," especially for transformative research in the middle and low ranges [6, 7, 25, 26]. But in crisis situations where something truly radical seems called for, even plausibility (as often understood) can be too restrictive.

[3] I have written on some of these matters before, e.g., [45], where I called confirmational appraisal *epistemic appraisal* (EA). That term conceded too much to the traditional view that epistemology begins in context of justification, whereas the very point of my work is to emphasize the epistemological relevance of research decisions and actions prior to the testing phase.

[4] Traditionalists are not committed to a completely static conception of confirmation. When I speak of theories, models, etc., already "on the table" I mean only ideas well developed enough to put to some sort of empirical test, not necessarily the proverbial "final product" ready for inclusion in textbooks.

3 The Frontier Problem Problem

When I was a naïve, young graduate student, I was a teaching assistant for Carl Hempel in a beginning philosophy of science course. In such courses it was then customary, when discussing the goals of science, to say that science is not merely concerned to find the truth about the world; it should be *interesting* truth. Today my (still naïve?) question is whether the emphasis should not be placed even more heavily on 'interesting' and correspondingly less on 'truth'.

The extreme case is exemplified by a (possibly apocryphal) story about Niels Bohr. Taking his customary walk with a visitor to his physics institute in Copenhagen, Bohr remarked to the younger physicist: "I think your ideas are crazy. The trouble is, they are not crazy enough!" Bohr's point was, of course, that physicists had worked themselves into a corner regarding the problem in question, and that a drastic change of theory would be necessary to escape.

Both logic and historiography of science tell us that most of our big theoretical ideas are likely false. In fact, as noted in passing above, most sciences deal more with models of various degrees of generality than with laws and theories. And in the case of models scientists usually know from the beginning that they are limited in some way as a representation of reality, since they involve idealization, simplification, approximation, and/or abstraction.[5] In other words, they are known to be false. So why all the philosophical emphasis on truth or verisimilitude, both of which are inaccessible to us when it comes to high-level claims about the universe? Here is where the question arises whether appeal to truth can furnish any additional methodological directive, any additional resources for research, a question that is still open. The relevant point here is: we do know that, in general, we are capable of raising and solving problems that we find interesting.

This is where heuristic appraisal (HA) comes in—as a guide to what is potentially both interesting and doable. Before saying more about HA, let me explain why I think it is important to inquiry in the most fundamental terms, going back to Plato's Meno paradox concerning the very possibility of genuine inquiry, inquiry that at least has a chance to be successful.

My problem of frontier problems is just a variation on Plato's Meno paradox. The well-known argument is that either we already know the solution to our problem (the answer to our question) or we do not. If we do already know, then we cannot genuinely inquire. And if we don't know, then we also cannot inquire, for then we would have no way to recognize the solution (answer) even should we stumble upon it accidentally. Thus, genuine inquiry (hence learning) is impossible. The problem is essentially a problem of recognizing something quite new as a possible solution.[6] The paradox, of course, is that inquiry is obviously possible, but how? The flaw in the Meno argument is that it places us in an all-or-nothing position regarding knowledge, failing to allow that there can be (fallible) cues as to whether

[5] See Shapere [66]. Levins [34] famously (and controversially) argued that in ecology one must sacrifice either realism, precision, or generality.

[6] Margolis stresses habits of mind and pattern recognition as the keys to cognition. See his [37, 38].

or not we are making progress toward a solution. In short, we can go between the horns of the dilemma. A few serious investigators (e.g., Herbert Simon [67]) have noted the importance of heuristics to the solution of the Meno problem, but the topic remains underdeveloped in general methodological and policy terms.

Heuristic appraisal (HA) is my label[7] for the identification and evaluation of hints and clues that can provide direction to inquiry in the sometimes large gap between the extremes of complete knowledge and complete ignorance. HA in its most general form thus concerns the generation or detection and evaluation of indicators at all levels, from an individual's effort to choose a good problem to work on (or to choose a good experimental design) to a high-level national funding committee deciding on its funding priorities.

Nearly everyone recognizes that heuristics, under some label, is a key to answering Meno. However, most writers drop the matter at this point, leaving hugely undeveloped how various forms of heuristic appraisal of candidate problems, solutions, and research techniques actually work. Given that "the problem of the growth of knowledge" (basically the Meno problem) is the central problem of epistemology, as Karl Popper rightly tells us [54, 55], it is surprising that philosophers of science, notoriously including Popper himself, have not paid more attention to these issues. As everyone knows, methodologists of science have given infinitely more attention to problems of justification than to problems of innovation at the frontiers of science. As noted above, the disciplinary focus remains on how to make decisions between or among already well-developed theories or models with reasonably extensive empirical and conceptual track records. There is little attempt to explain what motivated the generation of those candidates in the first place.

My own approach is the broadly Darwinian, naturalistic one of seeing all new design, including scientific and technological work, as emerging from iterated, hierarchical processes of trial and error, variation on extant forms and selection, with lookahead heuristics that become more specific and, in that respect, more powerful as apparent domain knowledge increases. As Edward Feigenbaum remarked, early in the history of artificial intelligence:

> There is a kind of 'law of nature' operating that relates problem solving generality (breadth of applicability) inversely to power (solution successes, efficiency, etc.) and power directly to specificity (task-specific information) [19, p. 167].

We get the same idea from Allen Newell:

> Evidently there is an inverse relationship between the generality of a method and its power. Each added condition in the problem statement is one more item that can be exploited in finding the solution, hence in increasing the power. [*Ibid*.]

For example, the General Problem Solver of Newell and Simon [42] was weak, because its heuristic rules were general, e.g., hill climbing and backward chaining. It incorporated no domain knowledge.

---

[7] Ernan McMullin first used this term, to my knowledge, in [39]. I must have "borrowed" it from him. I should add that he was a more thoroughgoing scientific realist than I am. He is not responsible for my excesses.

The "no free lunch" theorems of David Wolpert and William Macready [74–76; see also 46] claim that no search algorithm can be known to be better than any other, completely a priori, a point first broached by David Hume, whom they cite. Roughly speaking, when averaged across all possible domains, all search and optimization algorithms come out even. Thus a completely general method of science or of any other innovative activity, in the absence of domain knowledge already gained, is impossible, insofar as a method is something that concretely directs inquiry.

Strictly biological evolution possesses little or no lookahead power. The variation, selection, and transmission mechanisms are blind (undirected) to the future and the results (biological organisms) are directly field-tested, as it were. We humans possess somewhat more lookahead capability in terms of future planning and offline design and testing,[8] but it is quite limited, especially at frontiers of research. Articulating this lookahead capability in somewhat general terms, as informed by specific examples, is just what the study of HA, as I understand it, is about.

Since strong heuristics are domain- and content-specific, we should not expect to find a general account of HA that provides detailed direction to particular researches. However, this need not matter much for our purposes, because the scientific experts that evaluate proposals do possess the requisite domain skillset.

4 What Is Transformative Research?

There is transformative research of many kinds and degrees. Let us begin with kinds. There can be breakthroughs at any level of scientific activity—the reformulation of a key problem, the solution of such a problem, the invention of a new kind of instrumentation or mathematical or experimental technique, the organization of a new kind of laboratory or other organization, even the move from isolated research and development (R&D) departments to the democratization of innovation [3, 71]. One of the most important sorts of breakthroughs is the formulation of a new, previously unconceived or believed-unattainable *goal*, which generates the problem of how to achieve it. Given the interplay between goals and problems, it often happens in reverse: solving a problem or developing a new technique leads scientists to search for new goals that may now be within reach of the new capabilities. In effect, we have a solution in search of new problems. Another sort of breakthrough imposes a general constraint on future research. Think here of the conservation, invariance, and symmetry principles that transformed research in early 20th-century theoretical physics. Still another sort gives scientists far better instrumental access to a domain of phenomena than before.

In this chapter I shall be concerned with transformations in the sense of changes that challenge current understandings, either by undermining them or by opening up new areas of investigation that current views give us no reason to anticipate and that may even have been inconceivable before. I shall not be much concerned with breakthroughs in the sense of applications of already extant science and technology.

---

[8] See Dennett's "Tower of Generate and Test" in [13], pp. 373 ff.

This brings us to *degrees* of transformation. The highest degree is a full scientific revolution of a Kuhnian variety in which old understandings and practices are radically overturned and replaced [50]. Howard Margolis argues that Kuhnian revolutions such as those of Copernicus and Lavoisier amount to overcoming cognitive barriers (deeply ingrained habits of mind) rather than bridging gaps [37, 38]. There are surely other kinds of revolutions—in instrumentation or mathematical calculation techniques, for example. Peter Godfrey-Smith suggests that the extremely rapid progress in molecular biology is a "deluge" rather than a revolution of the Kuhn variety.[9]

5 How Identify Transformative Research in Advance?

I don't think it is possible realistically to plan (or fund) a successful revolution, and it is difficult to identify something as a revolution even while it is occurring, at least until it has been largely accomplished. Typically, what is accomplished is not what the instigators may originally have expected. The more profound the revolution, the more difficult it is to appreciate in advance the likely outcome and its far-reaching implications. Besides, 'revolution' is a success term, as opposed to 'attempted revolution', 'revolt', and 'rebellion'. Most attempts at revolutions fail, and those that do succeed typically develop in highly nonlinear ways. Often in science we can recognize that a revolution has occurred only in distant retrospect. We are wise only after the event. The nonlinearity point also follows from the fact that we sometimes find, retrospectively, that what triggered the revolution was a result or shift in practice or instrumentation that, at the time, seemed pretty normal rather than revolutionary. In such cases the eventual result of a seemingly ordinary cause is an enormous effect. The reverse happens as well: enormous efforts lead to little or northing. An example of the first is Planck's technically problematic derivation of his empirical black-body radiation law in 1900.[10] An example of the second is failure of the Newtonian tradition to find a mechanistic explanation for gravity. In sum, Popper is surely correct that "no society can predict, scientifically, its own future states of knowledge… [W]e cannot anticipate today what we shall know only tomorrow" [54, p. vii]. Our foresight is practically nil when it comes to predicting or forecasting scientific revolutions.

[9] In [22] Godfrey-Smith challenges Jablonka and Lamb's claims in [27] that evolutionary biology is now undergoing a Kuhnian revolution, on the ground that biology is not so tightly organized as physics. Jablonka and Lamb contend that evolutionary and developmental biology and the study of evolution of culture are now experiencing a multitracked revolution, a significant challenge to evolutionary orthodoxy in the name of evo-devo (evolutionary and developmental biology) and epigenetics. Epigenetics itself is clearly a transformative development, whether or not we should classify it as a Kuhnian revolution.

[10] See [31] for details. Examples are easily multiplied in all spheres of innovation. Who could have foreseen the import of 19th-century work on specific heats or spectral lines, or the invention of the transistor by Shockley, Bardeen, and Brattain in the late 1940s? Both have had immense impacts on both basic science and technological innovation.

Business analyst Peter Drucker once applied a military metaphor to describe the future that business firms face, in a way that nicely brings out the nonlinearities and discontinuities that have always made prediction of the future so uncertain:

> [T]he future is, of course, always "guerilla country" in which the unsuspected and apparently insignificant derail the massive and seemingly invincible trends of today… There will be discontinuities which, while still below the visible horizon, are already changing structure and meaning of economy, polity, and society. These discontinuities, rather than the massive momentum of the apparent trends, are likely to mold and shape our tomorrow… [16, p. ix].

The character of work at the frontier was captured inadvertently by U.S. Secretary of Defense Donald Rumsfeld when he said (in connection with his controversial management of the very controversial Iraq war) that there are known knowns, known unknowns, and still unknown unknowns. The unknown unknowns situation characterizes the more radical frontier research. As commentators on Rumsfeld's remark have noted, there are also unknown *knowns* in three quite different senses: things we think we know but really don't, things that are known (or have been known) by some but are unknown by others, and things that were once known (by the same of different people) but that are now forgotten.

Finally, Hans-Jörg Rheinberger conveys what frontier life is like in scientific research when he writes from experience that

> the experimental scientist deals with systems of experiments that usually are not well defined and do not provide clear answers. […] Experimental systems are to be seen as the smallest integral working units of research. As such, they are systems of manipulation designed to give unknown answers to questions that the experimenters themselves are not yet able clearly to ask. Such setups are, as Jacob once put it, "machines for making the future." They are not simply experimental devices that generate answers; experimental systems are vehicles for materializing questions… [61, pp. 27 f]…
>
> Experimental systems, together with the scientific objects wrapped up in them, are inherently open, if bottlenecked, arrangements. Their movement is such that it cannot be predicted if they are to retain their character as research devices. Epistemic things, let alone their eventual transformation into technical objects and vice versa, usually cannot be anticipated when an experimental arrangement is taking shape. But once a surprising result has emerged, has proved to be more than of an ephemeral character, and has been sufficiently stabilized, it becomes more and more difficult, even for the participants, to avoid the illusion that it is the inevitable product of a logical inquiry or of a teleology of the experimental process. [*Ibid.*, p. 75].

Now none of this is to say that a bold scientist might not find reason try to foment a scientific rebellion. However, if we think of policies that funding agencies might adopt, it seems foolish to suppose that there could be a workable revolution-identifying policy. Instead, I shall leave Kuhnian revolutions to one side and focus on what we might term transformative science "of the middle range" (with a bow to Robert Merton [41]).

Even with this much grasp of what transformative research is, it is not easy to *identify* promising, potentially transformative research; for, again, what is transformative is often apparent only in retrospect. The main reasons are those mentioned above. In any highly creative enterprise the misses will typically far outnumber the hits, and scientific advance can be highly nonlinear.

I also suggested above that we don't need a detailed account of HA in order to proceed with policy reform, since, at crucial stages of the decision-making process, domain experts skilled in such matters will be involved. But here a related problem arises, for how are they to do their job of HA? How is HA to work in contexts in which we seek transformative research? For insofar as HA is tied to current domain knowledge, HA can itself be conservative, something akin to the HA practiced by Kuhnian normal scientists. Yet insofar as it is *not* tied to current domain knowledge, HA becomes weaker in its search-directing power.

There is a double recognition problem here, a double Meno problem. A transformative proposal (or, for that matter, work already done and published) must be recognized as both a serious contribution to the field and yet as transforming the field. The expert community is increasingly less likely to recognize departures from orthodoxy as serious contributions to the field insofar as they are radical. And work that would turn out to be transformative is difficult to recognize as such, unless it *is* clearly radical. So, we have:

> **The essential tension.** Insofar as a recognizably serious line of research would be transformative, it is difficult to recognize and to accept as such; and insofar as it is easy to recognize, it is normal science and not transformative.

Kuhn dubbed this phenomenon "the essential tension" between tradition and innovation [30]. In decision-theory terms, it places decision makers in the regime of decision-making under uncertainty already at the problem-recognition level. And the uncertainty increases rapidly with degree of transformative potential. There are degrees of uncertainty.

Don't get me wrong. Something like systematic normal science is crucial to scientific progress, although Kuhnian normal science is widely agreed to be too dogmatic. At the opposite extreme it is easy to agree with Kuhn that Popper's idea of "revolution in perpetuity" through severe criticism of fundamentals would be incoherent and a bad science policy to promote, at least if we understand scientific revolution in anything like Kuhn's sense. The problem is to find an appropriate balance between Popper and Kuhn. "Transformative research of the middle range" is my placeholder for this balance. In economic terms, a related source of balance is division of labor within a specialist community, where some members are engaged mainly in exploitation (normal science), while others are engaged in exploration [11, 12, 64, 65].

6 The Special Problem of Transformative HA: Summing Up the Argument

This section is a partial articulation of subproblem 1, mentioned in Sect. 2: How do scientists and supporting institutions make fertility decisions at frontiers?

1. The No Free Lunch theorems state (roughly) that no method, no search algorithm, works better than any other when averaged across all possible domains.
2. Thus there is no such thing as a general, a priori scientific method that concretely directs inquiry. There is no general, a priori logic of discovery, nor an a priori logic of justification either.

3. Thus, insofar as a domain is new, we cannot know in advance whether any method will work, or which method will work better than others. How precisely to proceed presents scientists with decision problems under uncertainty.
4. Strong heuristics require substantial domain knowledge, well-structured search spaces.
5. Transformative research contexts (especially of the kind that we are focusing on here—those that are disruptive rather than simply rapidly additive) present new domains (or domain characterizations) that are not yet well structured.[11]
6. Hence, strong heuristics tend not to be available in the initial stages of transformative research.
7. Hence, the choice of what method (including which heuristics) to try is itself a weakly-directed choice among weak methods. It is itself an instance of nearly blind variation and selective retention, for at this point, given the absence of constraints, even the choice of which method to try is blind [5, 13, 44, 53].
8. Decisions made at transformative frontiers, by definition, have a greater potential impact on the direction of science than normal scientific decisions.
9. Thus our search heuristics are weak in rough proportion to both the importance and the uncertainty of our decisions. The more we need HA, the weaker it usually is!
10. Thus we have little foresight or control over the potentially deepest innovations.

7 The Difficulty of Transformative Science Policy: The NSF and Lisbon 2000 Examples

Here I begin articulating, in somewhat practical terms, our second subproblem 2: *Which factors stimulate or hinder transformative frontier opportunities?*

Given that transformative innovation, by definition, results in faster progress than incremental innovation, there is pressure on funding institutions to support potentially transformative research, despite the increased risk and uncertainty. However, as we have seen, there is little agreement at this level on what potentially transformative research is or on how to recognize it. For by its nature it will often be something "different," something unexpected. Moreover, there are counter-pressures toward conservatism in both democratic government and business enterprises. Taxpayer and profitability concerns demand visible payoff in the short run. There is also the fear of embarrassment of too quickly supporting projects that are defective or simply bogus, as in the Utah state government's rush to celebrate the alleged discovery of table-top cold fusion. Many analysts have pointed out that the

[11] Some qualification is needed here and at other points. To some degree, strong nonlnearity undermines Kuhn's normal-revolutionary distinction. Since even apparently normal work can trigger a transformative change if the state of the field is just right (the nonlinearity point made earlier), work at this stage can still be strongly directed. The crunch comes in the positive work that aims to re-orient the field.

quality-control filters almost necessarily impose a conservative bias on the grants process. The issue is, how much quality control is too much? What is a healthy balance between "variation" and "selection"?

Here I shall focus on the U.S. National Science Foundation (NSF), then add a few comments about similar problems in the European Union. The majority of NSF grants are to individual projects. In most NSF programs proposals are first vetted by at least three external expert reviewers and then by panels of experts who make recommendations to higher-level program and division officers, who balance the overall grant portfolio in various ways. At its fastest, this process takes many months. When revisions are required over several granting cycles, it can take years. Since there is stiff competition for scarce funds, nearly any significant defect or oversight in a proposal (even in how the data will be secured or shared) is enough to kill it for that round. Although resubmissions are possible, killing a promising proposal for a couple of rounds, sometimes on the basis of bureaucratic technicalities, slows the proposed research process. Committee consensus is the basic decision process on at least one level, and the individual committee members wish to appear rigorous and tough-minded to their peers. Thus panelists are naturally inclined to look for reasons to reject. Committee decisions tend to be conservative compromises in any case. Given the constraints, the present system is too often an attractor for critical rejection of proposals rather than one identifying potential breakthroughs.

To counter the perceived conservatism, NSF now places a premium on proposals judged to be "transformative." However, it also has imposed a "broader impacts" criterion that is sometimes interpreted to mean that good proposals should lead either to practical results in the near future or (at the very least) should be obviously useful to other fields of academic research.[12] So interpreted, these two criteria will often be in tension. Genuinely creative breakthroughs, as we know from the history of science, are typically quite technical, i.e., initially specific to a particular subspecialty. It may take decades to work out the larger implications. Who could have known that the invention of the transistor would ultimately transform scientific modeling, search, and computation across the board and generate whole new specialty areas?

To be fair, there is now wide discussion of the problem of identifying and funding transformative research. It is being addressed explicitly inside and outside of granting agencies in the European Union, the USA, and around the world. At the annual meeting of the American Association for the Advancement of Science in February 2013, for example, several sessions were devoted to the problem. One of these was a report on similar difficulties in the U.K. and the European Union. For example, the Lisbon 2000 Agenda was to give the EU a competitive advantage in technoscience by 2010, based on increasing the gross domestic product (GDP) investment in science to 3 % from approximately 1 %—while at the same time making technoscience more responsive to public concerns. While everyone wants

[12] This is an odd interpretation, since NSF was founded to support basic research, not to engage in the applied activities of any number of corporations, government agencies, and non-governmental organizations. Thanks to Kelly Moore for this point.

scientific research to be socially beneficial, the EU's social constraints on the process could only introduce still more layers of red tape [10, 48, 49, 70]. Needless to say, the goals were not achieved and would not have been achieved even without the economic crash that began on Wall Street. The EU techno-economic mandate itself demanded just the sort of short-term accounting that is unlikely to lead to efficient grants processes and sustainable results.

I shall sum up the problem this way:

> **The policy problem**: Most funding agencies (especially in government) are structured to discourage transformative HA recognition and/or (once recognized) to undervalue it in the interests of short-term accounting. History informs us that the innovation timescale is typically an order of magnitude or more larger than the de facto accounting timescale imposed by such requirements as "broader impacts." There is too much risk-avoidance, too much emphasis on quasi-guaranteed results.

8 Traditional Philosophical Approaches: CA Versus HA?

Traditional philosophy has not helped. First it excluded context of discovery from epistemology altogether, on the ground that only context of justification was of rational, epistemic interest. More recently, some philosophers have discussed topics that fall under my umbrella term, "heuristic appraisal." However, many of them, in effect, attempt to reduce HA to some version of confirmational appraisal (CA), often in a somewhat weakened form so as not to reject new initiatives too quickly.

There is a strong temptation to do this, because it seems perfectly legitimate, perfectly rational. The justificativon runs like this. "Research problems are set by goals and standards that must be met if eventual success is to be achieved. So it is fair to evaluate the progress of the new research with respect to those goals. In fact, such evaluation is necessary to determine which problems still need to be addressed, and in which order."

My response: transformative research is typically a multistage, evolutionary process. True, the initial formulations of the problems are conceived in this way. However, as materials cited above indicate, most creative research involves goal shifts and hence problem shifts, and transformative research does this to a greater degree and can require the construction of new standards for the new domain that may be opening up. These shifts are no small matter, for, again, the sort of HA that is most directive at transformational frontiers is highly context-and problem-sensitive.

Initially, it was worth considering the heuristic that promising new theories were the ones that would reduce to the old ones that worked well under some limit or approximation (supposedly thereby explaining *why* they worked so well). Popper long touted this idea, somewhat indebted to Bohr's correspondence principle. In [56] Heinz Post developed it in interesting detail. While certainly a useful way to vary current exemplars and to check the results for plausibility, the heuristic is surely too strong, too conservative, to raise to the level of a justificatory requirement, for intertheory/intermodal relations can be too complex for that. Quantum mechanics is more than a variation on classical mechanics. In fact, 'in praise of conservative induction' was Post's subtitle. A related tendency is for philosophers to require that later theories "cohere" in some more general manner with earlier ones and with relevant background knowledge, in an overly conservative sense of coherence.

Another shortcoming, mentioned above, is to ignore the early stages of development by assuming that the new approach is already fairly well developed and ready for head-to-head comparison. For example, Larry Laudan [33] attempted to address the problem of the "pursuit" of new theories by comparing the rate of problem-solving progress rather than cumulative progress to date, in order to explain the epistemic attraction of new initiatives. His work was insightful. However, he supposed that semi-mature comparisons were already possible. While a step in the right direction, this, to my mind, does not fully address the problem of how such lines of research are motivated in the first place.

In *Structure* Kuhn visibly raised a major problem for traditional CA, namely, the problem of new theories.[13] The problem is to explain why scientists in their right mind would abandon a well-established theory or paradigm (or deeply ingrained set of practices) for one exists only as a crude sketch of a research program. How could such an endeavor hope to go head-to-head with a successfully established theory? Although Kuhn formulated the problem in the light of his conception of revolutionary overturning, the problem arises, to an interesting degree, even when a maverick group proposes a new direction to the field that is compatible with the old framework, yet unwelcome for some reason. Perhaps the technical ability required versus the perceived payoff is the issue, as in late 19th-century statistical mechanics and today's string theory. The counterpart in business would be for a company to introduce a new product line that competed, without completely displacing, a successful line that it already has (see [8]).

In effect, Kuhn contended that the problem of new theories was an artifact of the philosopher's retrospective view of scientific work. HA was the key to Kuhn's own attempt to solve the problem of new theories or new paradigms. As is well known he rejected traditional confirmation theory. And although he also rejected the idea of a logic of discovery (and that of rule-based methodologies more generally), his aim in these parts of *Structure* was to explore decision-making in context of discovery. First he pointed out that "discoveries" are *articulated*, in terms of both their development by the original investigator(s) and their critical discussion and further development by the community. They are not instantaneous "aha!" experiences. Then he stressed that future promise, expected fertility, is a strong motivator of scientists' research commitments. His point seems to have a Cartesian existential dimension: one retains one's self-identity as a creative scientist only as long as one is actively working on important and challenging problems. Scientists exist as creative scientists only as long as they are actively thinking scientifically about such problems. Merely teaching the subject does not count, and completely

---

[13] Bas van Fraassen [69, p. 125] both addresses the problem of new theories and sharply rejects traditional confirmation theory: "An almost century-long, practically fruitless effort to codify evidential relations (so-called confirmation theory, a bit of bombast if anything is) should have convinced us that 'in accord with experience' is not a simple, uncritically usable notion."

routine work counts for little. Thus even a warhorse of an old theory will become a less attractive research site once it is thought to be "mined out," with few resources left to respond either to old anomalies or to generate new goals and new problems. In these cases HA can trump CA. The overall motivation of normal science itself, Kuhn told us, employing Aristotelian language, is to actualize the potentiality of a new paradigm [29, p. 24]. But once actualized, the potentiality is gone.

In a rich, recent paper [23], Leah Henderson and colleagues address the problem of new theories directly, from a hierarchical Bayesian confirmational point of view. Theirs is a sophisticated attempt to get beyond single-level Bayesian inference, applied to the successive testing of a hypothesis, in order to consider rational theory change and the problem of new theories. The authors make several excellent points, e.g., confirmation of change at one level can appeal to higher theoretical levels as well as to empirical test results. Frontier research can be guided by higher-level theories that are well confirmed. This is certainly true in some cases, but what if the change is at the highest level and there is no higher level to justify it? In this respect their paper falls victim to a finite regress, failing precisely where a Bayesian needs it most. Their overall effort amounts, once again, to attempting to reduce HA to CA.[14] Pure bayesianism itself is, of course, one approach to mathematical probability theory, a content-neutral formal calculus. Induction from the data plus a few theoretical constraints from the background knowledge will rarely provide much guidance in transformative contexts. By contrast, in generative contexts HA employs rhetorical tropes that are laden with material scientific content.[15]

9 Diagnosis of the Short-Term Accounting Problem and the Suppression of HA

What contributes to the neglect of HA by philosophers and also by those organizations that have, usually unwittingly, erected obstacles to the full expression of HA? In [45] I identified some disciplinary factors, including hard-headed philosophers' love of precise calculi. HA looks quite messy by comparison. Here are some additional, more or less philosophical considerations.

Economic motives. Perhaps the most obvious consideration is that no one wants to waste resources on failures. To taxpayers, legislators, and CEOs, variation-selection processes look incredibly wasteful, whether they are biological (e.g., trees or fish

[14] There are several difficulties with their sort of position, in my opinion, besides the usual ones specific to Bayesian approaches [6, 7, 21, 25, 26, 68]. Given that they are not naïve empiricists, it is surprising that empirical curve fitting dominates their examples of breakthrough research and theory dynamics. Their HA component is almost entirely past-directed to what is already "on the table," thus reducing HA to CA. And they take orthodox science to be stable and robust, not fragile.

[15] Subjective Bayesians will respond that all of the HA considerations can be included in the prior probabilities, but this move leaves HA in the traditional domain of individual psychology, without providing further analysis.

emitting thousands or millions of seeds or eggs, only a few of which will grow to reproductive age) or whether they occur in scientific and technological research, where the tax dollars are being spent. To these groups, long-term investment in an uncertain future seems like a poor investment. (Future discounting issues also come into play.) The only remedy that I can think of is helping the relevant parties to understand more clearly the way creative enterprises work, spiced by scintillating stories of developments they can appreciate. In the longer term this means more focus in school (as well as in philosophy!) on what Bruno Latour, in *Science in Action*, calls science and technology "in the making" in place of science "ready made," taught as a collection of factoids [32, p. 4].

Political and journalistic rhetoric. Reinforced by commercial journalism's goal to create controversy in order to boost sales, politicians embrace the culture of blame—"the blame game." Failure to make rapid scientific and/or technological progress despite large investment is regarded as a failure in judgment or competence for which political and moral as well as scientific blame is to be assigned. Nor do researchers like to be embarrassed by being accused of incompetence or graft when things don't work out. This is another dimension of the lack of understanding of decision-making under risk and under uncertainty.

Intelligent design. Legislators and the general public they serve too often remain committed to an "intelligent design" conception of human creative innovation in general, and especially in the sciences. The most common expression of this idea is that there exists a universal "scientific method," by the boring, tedious application of which worthwhile results should be forthcoming. As if there could be a step-by-step method, known in advance, for negotiating unknown territory! Good methods for a novel domain are typically the end-result of inquiry, not the beginning. Only at the end of successful inquiry are frontier problems tamed to the point that they become routinely solvable. Only then can powerful methods that incorporate substantial domain knowledge be formulated. While it is true that we humans have more lookahead and off-line testing capabilities than the rest of known biological nature, this ability is quite limited.

The end-of-history fallacy. This is my label for the common mistake of thinking that we stand at or near the end of history in the sense that major historical transformations in a supposedly mature science are all in the past, that the future will be a flat extension of the present. In the case of basic science this means no more scientific revolutions, no more truly major breakthroughs in such fields. (A more moderate version of the end-of-history thesis is that transformative breakthroughs will become successively smaller as we converge on the truth about the natural world.) This view often seems to be coupled with a second sense of 'the end of history', namely, the idea that scientific research has matured to the point that the historical-cultural horizons that limited previous work have now been removed so that current results stand outside of history. This is the view that strong predictive and explanatory confirmation is sufficient to detach scientific claims from historical path dependence. Third, this in turn supports the view, associated with Cartesian-Enlightenment reason, that we can survey all logical possibilities and impossibilities a priori.

The issues here are difficult and hotly contested. As indicated above, my own position respects Hesse's principle of no historical privilege. As Carlo Cellucci reminds us, in crucial respects, we are just as historically located as was Aristotle.

While we can legitimately claim progress beyond our scientific predecessors, when it comes to frontier research we are in essentially the same situation as they were. To be sure, we now have more, and more powerful, research tools than before, but those tools have opened up entirely new frontiers that they may or may not be capable of mapping. Our own horizons of language, reason, and imagination must remain largely invisible to us at any time, when (like the objects in the rear-view mirror of a vehicle) they may in fact be closer to us and larger than we think. In this respect, ironically, deep historical studies have helped to make this problem worse rather than better, since we can have no deep history of the future. Via the non-privilege principle we can have only the pale supposition that it may be equally deep.

In general we cannot today say what current work will be considered most significant in the future, for historical significance is just that. The "history" (the historical evaluation) has not yet happened for today's work. The wisdom of historical significance takes flight at dusk. Nietzsche makes this Hegelian point in "On the Uses and Disadvantages of History for Life" [52, §2], and it had been made previously, quite unintentionally, by the president of the Linnaean Society (where the first papers on evolution by Darwin and Wallace had been presented), who stated in his annual report for 1859 that nothing of transformative and enduring significance had happened during that year.[16]

In my opinion Hesse, Kuhn, and Paul Feyerabend were the first major historians and philosophers of science to avoid the end-of-history fallacy.

Caution is needed here, because creative scientists themselves justifiably conclude in some cases that the main problems of a domain have been solved and that it is time to move on to greener pastures. In short, there are ways of making justified assessments of the degree of "wildness" of a frontier of research. While these judgments remain historically relative (consider judgments of the state of Euclidean geometry in 1800), fertility considerations combine with problem-solving and predictive success to establish their importance. Naturally, it is a bad idea for funding agencies to pour resources into areas that seem particularly unpromising of new results, no matter how productive those areas have been in the past.

Overly strong scientific realism. An interesting question that I cannot pursue here is whether there is a strong form of scientific realism that evades the end-of-history fallacy, that is, a form of realism that claims that we are already near the truth, in apparent violation of Hesse's principle. The cautionary note of the previous paragraph signals that a local, "spotty" realism may be more defensible, where the spottiness may be temporal as well as distributed over the scientific specialties represented in a given time-slice. Legitimate claims of realism can be rescinded if

[16] Cohen [9, p. 286] provides the key quotation: It is "only at remote intervals that we can reasonably expect any sudden and brilliant innovation which shall produce a marked and permanent impress on the character of any branch of knowledge." The appearance of a "Bacon or a Newton, an Oersted or a Wheatstone, a Davy or a Daguerre, is an occasional phenomenon whose existence and career seem to be specially appointed by Providence, for the purpose of effecting some great important change in the conditions or pursuits of man." [His main point was] that the "year which has passed has not, indeed, been marked by any of those striking discoveries which at once revolutionize, so to speak, the department of science on which they bear." Thanks to Devin Bray for the Nietzsche reference.

and when a domain gets re-opened by transformative work. Nonetheless, I believe that strong realist commitments provide a bias against transformative research proposals of the sort that undermines present understandings.

Popper always argued that realism is superior to instrumentalism in stimulating further research, as if instrumentalists don't really worry about precision. My present point is not to defend classical instrumentalism but instead to underline the claim that strong realism can also dampen expectations for transformative research. Popper's realism was a form of weak, semantic realism, one that says only that theoretical terms are meaningful, hence that theoretical statements make true or false claims about the world. Popper treated all universal lawlike claims as conjectures with zero probability, so he obviously did not endorse the strong epistemological realism of people such as Richard Boyd and Stathis Psillos that claims, in addition, that our present, mature theories are close to the truth [4, 57].

It is this strong realism, not semantic realism, that worries me when transformative frontier research is in question. The "no miracles argument" for strong realism is that the many great successes of current, mature sciences would have negligible probability if these sciences were not close to the truth. But what if we frame our historical situation somewhat differently so as to bring out our future vulnerability? "Given our long history of success, we must be fairly close to the truth now, so it would be a miracle if major departures from present orthodoxies were to succeed." The strong realists are hostages to fortune in predicting (in effect) the nonexistence of major future revolutions and, to a lesser degree, significant transformations that seriously alter current understandings. And this perhaps false sense of security, I fear, biases evaluation in favor of CA over HA.

The claim that transformative change is becoming less severe as we converge on the truth is supported by neither logic nor historical time-series analysis, as far as I can tell. And if something like Kuhn's view in *Structure* is correct, then we may expect some future revolutions to be larger rather than successively smaller [50]. One suspects that strong realists are incomplete fallibilists, retaining a whiff of the foundationist, truth-preserving impulse (compare [7] on mathematics).

Philosophers' tendency to optimize or maximize in setting goals and standards. Claims of final truth, greatest good, perfect beings, etc., fit real world practices badly and are often too "out there" to serve even as regulatory ideals. This overreaching in turn inspires hypercriticism when the enterprises fall short of the idealized goals. The perfect is made the enemy of the good, the best the enemy of the satisfactory.

Philosophers' over-concern with logic and rationality and with wrong kinds of coherence. Logical inconsistency is an extreme case of incoherence, the worst sin for the logical empiricists and an immediate "knock off" of a new proposal. Yet the history of transformative science is remarkably inconsistency-tolerant [40]. More moderately, Popperian correspondence and other strong forms of coherence with the past (e.g., via limit relationships) contain a residue of epistemic foundationism, based on the idea that deductive logic and even probability theory (to a lesser degree) give us a knowledge-preserving ratchet.[17]

[17] See [7, 20, 25]. On diversity-tolerant coherence see [62, 63].

10 What Is to Be Done? and What Role for HA?

It is easy to make the negative case against traditional confirmation theory's claim to be the key to decision-making at research frontiers. It is more difficult to make the positive case for HA in terms that are reasonably crisp. The parallel problem is evident in the dozens of self-help books in the business and economics section of bookstores—books on how to creatively transform your business. Such books typically contain, or mention, a few case studies but then become pretty vague when attempting to generalize or to project onto the future.[18]

To some degree this cannot be helped. For on my own view there can be no answer to the question of this section that is both general and that also provides focused, short-term direction across the spectrum of frontier research domains. Finding general, content-free rules is a non-starter. One has to proceed on a case-by-case basis, bringing expert knowledge into play, but in the right sort of (non-conservative) manner. Case-based and model-based reasoning [36] can work very well in specific domains, since the cases can provide fruitful exemplars, guiding precedents.

Since we are looking for general policy advice for the longer term, it suffices to focus on removing barriers and creating general opportunities rather than on pretending to give specific directions to the specialists in their domains. Accordingly, I shall lay out some fairly general suggestions that may work in a number of cases to generate expert discussion of possibilities. My point here is that explicit and implicit philosophical realist and administrative biases are often barriers to fully developed HA. In short, these biases privilege current CA over HA. Unfortunately, some of these biases would be difficult to remove, since they derive from public accountability requirements of both government and corporate funding agencies.

The following argument extends that of Sect. 6.

1. Universal BVSR: All creative innovation is the product of (nested) blind-variation-plus-selective-retention processes. Thus one must expect frequent failures in novel domains, contrary to realists and others who argue that our improved methods and instruments should increase our rate of success over that of our predecessors.[19]
2. Therefore, there is no general, content-neutral scientific methodology.
   Comment: The BVSR claim seems to be the main route by which Popper and Campbell arrive at this conclusion, which also connects with the later work of Wolpert and Macready cited above. I agree that the negative conclusion is correct. However, this cloud has a silver lining relevant to policy, for…

---

[18] The history of economics and that of philosophy of science are somewhat parallel in leaving innovation out of "theory," relegating it to "outsider" status as an exogenous factor that occasionally disturbs the equilibrium of the system. See [2, 73] as well as [50].

[19] I believe that [14] commits this mistake. The frontiers of our predecessors are usually no longer our frontiers. We have new frontiers, unimagined by our predecessors. The new dispensation may dramatically reorganize the search spaces. For example, today's epigenetics research largely overturns the old nature-nurture distinction. Search spaces increase exponentially in size as we realize that thousands of genes and epigenetic factors seem to be involved in some phenotypic traits. Ditto for neural networks and their embodiment.

3. Evolutionary computation is a family of BVSR problem-solving tools (e.g., genetic algorithms) that constitute the beginning of a semi-mechanizable scaling-up of search and discovery. These tools are variations on the so-called method of hypothesis (itself an instance of variation-selection) but are in fact more methodical by providing a generative component that was notoriously missing from the traditional account. These tools scale up the old practice by orders of magnitude, in principle and, increasingly, in practice. So this is no mere intellectual dream. These tools or methods are already employed in hundreds if not thousands of technical papers published each year. Think, for example, of the semi-automated methods that pharmaceutical companies use to search for bioactive compounds in hundreds or thousands of compounds at a time. To be sure, we are still far from having machines that can invent highly sophisticated novel theories, but that objection misses the present point—that our human methods are really just a slow form of BVSR in which we develop and test one or a few hypotheses or models at a time. The semi-automation already achieved is enough to show that variation-selection (at bottom, trial and error) should be considered a method and not the complete absence of method [46].
4. So universal BVSR does have methodological—and policy—import, but at a high level of abstraction. It provides a methodological "shell" akin to many other mathematical/computational tools. It has methodological "bite" but not a specific bite. To make it more specific, domain and problem-level constraints must be added. (A limitation is that current implementation still involves merely "mechanical" recombination, whereas human creativity depends heavily on modeling on exemplars employing analogy, metaphor, etc. [36]). It also has educational bite in helping people to realize that frontier research is very much a trial-and-error affair. The mechanization makes it possible to speak of discovery algorithms, in the broad sense, but not in the sense of algorithms that are bound to succeed at all, let alone on the first—or the thousandth—try (cf. [13]). The difference between trial and error (once one's guiding heuristics run out) and step-by-step procedure is one of scale and degree of systematicity, not an intrinsic difference of method versus non-method.
5. Hence my present position: General policies to enhance research should be pitched at this level as well as at transformative research of the middle range. At this level policy is concerned with infrastructure, with the goal of increasing opportunities by removing obstacles rather than that of legislating specific positive directions for research. A second goal is better education about the nature of work at frontiers and the corresponding development of what might be called a frontier epistemology.

Notice that our policy-making itself is a BVSR process, since we are also in a frontier situation at the policy level. So we begin by applying HA at the general policy level itself. We ask which ways of organizing the research effort, including support, are more likely than others to stimulate transformative research. Within grants organizations the question becomes how to achieve in practice greater transformative HA weightings. More HA specificity and hence policy specificity then naturally come into play as we descend toward more specific areas of science and technology, with their content-laden and problem-specific heuristics. Where taxpayers supply the funding, some attention to national and regional needs is necessary, but weighting these too heavily can be at cross-purposes with scientific judgments about which research areas are pregnant with possibilities. This, of course, is part of the overall problem, since politicians want to address the perceived current needs of their constituents.

Some general policy points (overlapping). Obvious suggestions include increased funding for research and better science education of both future scientists and of the general public. However, my concern here is what can be done short of this "brute force" measure, as major increases in funding are unlikely. What is more doable is to remove infrastructure barriers to research in general, including long-term innovation. In the USA and worldwide there has been a long-term immigration of basic research from the private sector and large government facilities to universities. This movement is increasingly true of translational ("applied") research as well. This despite the noted inefficiency of getting research proposals through the various state university and funding hierarchies. In the U.S., for example, state universities cannot compete with the MIT model. MIT can offer greater freedom to faculty members to found and run their own companies on the side. For government regulation is unavoidably heavier for state than for private universities given the multiplicity of governments, from federal to local, that have regulatory authority. Still, significant improvement is possible, especially among institutions still undergoing transition to first-rate research universities.

A second, more specific priority is also obvious and is already being implemented in both academe and industry, namely, to focus on a limited clusters of related specialties, instead of trying to be *Grands Magasins* of research. There are calls for collaboration across departmental lines, especially within the clusters. Just as the more innovative companies have tended to break down internal departmental boundaries, so universities are encouraged to do so—and more. The move from interdisciplinary to transdisciplinary research is Goal 1 of the American Academy of Arts and Sciences 2013 (see [1]).

There are several models for ramping up research activity in general, and some for encouraging transformative research. I briefly mention a few, then give a bit more attention to scenario planning.

The prizes/awards model. The idea here is to depart from general awards for work already done, such as the Nobel Prize and the Fields Medal (mathematics), and to return to the old practice of having prize competitions for specific problems.

The Linus Pauling model is my label for the Nobel laureate's proposal that a certain percentage of federal research funds be set aside for the use of a select few highly creative scientists with proven records of highly fertile research, with a minimum of advance accountability required in terms of the rigorous vetting of proposals. This policy would surely be difficult for democratic governments to defend.

The NSF model was discussed above with its potential conflicts between transformative proposals and broader impacts. The apparent conflict with the transformative priority needs to be removed. NSF was founded to support basic research, not translational research. Many other organizations, private, public, and nongovernmental do the latter (for more detail see my [46]).

The DARPA model (U.S. Defense Advanced Projects Agency) brings together key university and corporate researchers, including scientists, engineers, and business people, for highly targeted projects with a completion time of perhaps 2 years. Administratively, DARPA is small, flat, and flexible. The average project manager remains on the job for only about 4 years [15]. An interesting feature of this model is that DARPA is an agency within the U.S. federal government (one focused on perceived national security needs). Yet, despite its support by taxpayers, it avoids the time consuming, stage models of proposal writing and vetting that NSF and the National Institutes of Health (NIH) normally practice, not to mention university research support. This model works well for focused, translational research that requires no major transformation in basic science, but, for those reasons, is not a good model for longer-term, more basic research.

The "triple helix" model integrates university, private industry, and government resources in order to promote focused innovative research [18]. The goal here is to overcome the traditional, basic research orientation of universities, in which "going commercial" has not been a priority and has even been discouraged. The result is that the excellent research being done in universities languishes in the technical journals unless and until an unrelated commercial enterprise happens to spot a new development as reported in a journal article and to decide to develop it. This model, in its various forms of implementation, is what many universities are now adopting. The rapid reduction of state support for (so-called) state universities in the USA is indirectly forcing institutions to consider such a model in order to increase revenues. The triple helix model is three-way interactive. Instead of a company developing a university finding and perhaps hiring key university researchers part-time as consultants, university experts can also reach out to companies and to governments, asking them what they need, what problems need solving.[20] There are dangers here, of course, one being the erosion of basic research; another being a change in university culture as the liberal arts lose status.

The triple helix model has, of course, been scaled up. One prominent example is Stanford University with SRI International next door. Originally Stanford Research Institute, a center for innovation, SRI is now an independent contractor and is emulated by other institutions around the world. Still bigger, of course, is Silicon Valley as a whole, which originally was modeled on MIT and the Boston area, with the cluster of professor-founded companies along Route 128 (see [62]). On somewhat the same scale are the various "science cities" around the world, such as Tsukuba, Japan, where a high percentage of nationally-supported research is done.

The Rockefeller Foundation model focuses support on an area deemed by heuristic appraisal to be ripe for rapid, interconnected development. The foundation provided heavy support for research and scholarships in the biomedical sciences. Here is Lily Kay on high-level HA:

---

[20] Here I am indebted to Harvey Wagner. For a variation on this idea, see his [72] for a description of the Teknekron business plan.

> During the 1930s a new biology came into being that by the late 1950s was to endow scientists with unprecedented power over life… The aim of this book is to understand the historical process that propelled molecular biology to its dominant disciplinary status by uncovering the motivations and mechanisms empowering its ascent… [T]he Rockefeller Foundation served as the principal patron of molecular biology from the 1930s to the 1950s; Caltech, a primary site for implementing the Foundation's project, became the most influential international center for research and training in molecular biology. Why did the Rockefeller Foundation launch and sustain with massive support a new biology program at that moment in history? From the entire range of contemporary biological interests, why did scientists and their patrons privilege and promote a molecular study of life? Why was Caltech selected as a primary site…? [28, pp. 3f].

Universities that focus on identified research concentrations do this sort of thing on a smaller scale.

11 The Scenario Planning Idea

It was originally the idea of Herman Kahn at RAND Corporation and, later, of Pierre Wack of Royal Dutch Shell Oil Company to raise the company's future prospects by creating a few scenarios of possible futures, each of which challenged some established "truth" widely expected to hold in the future. (In this respect the strategy fits our displacement conception of transformation rather than simply the rapid amplification conception). One of these assumptions was that the huge energy resources within the borders of the Soviet Union would remain forever off-limits to Western corporations. But what if this common assumption were mistaken? Challenging this assumption was the basis of one of Shell's scenarios. A small group of planners kept their eyes on an unorthodox but rising Soviet economist named Gorbachev, eventually to become head of state. As a result, when the almost unbelievable happened and the Soviet Union collapsed, Shell was ahead of its competition.

Since much has been written about scenario planning since then, it suffices here to mention that expert scenario planners do not attempt to cover all possibilities, whatever that could mean (see, e.g., [35, 58]). Instead, they develop in some detail perhaps three or four scenarios distributed over different areas of concern—financial, technological, political, etc. Nor are scenarios something to which a probability is assigned. None of them are probable in any realistically measurable sense. The strategy is to keep the planners flexible—on the alert for hints of meaningful changes—not only with regard to these scenarios but also more generally.

If scenario planning works in the business world among firms competing on the treacherous terrain that is the future, why not try it for science policy as well? I have already noted some relevant parallels between scientific innovation and business innovation. Given that NSF, NIH, and other national granting agencies already convene workshops on the future of various disciplines, something like scenario planning is surely already being done, to some degree, on these special occasions. (In facing problems such as quantum gravity, for example, it is pretty clear that significant assumptions must be challenged.) If so, I suggest a wider reach for such efforts. How about funding special initiatives in key areas in which researchers are invited to challenge orthodox practices in a more significant way than in the normal research proposal? This would seem especially appropriate where an established result appeared to block otherwise fruitful approaches. The diffusion of this idea beyond special programs would likely help to shake up the normal grants procedure a bit, to inject a bit of "chaos" into an overly rigid system.

Naturally, such a policy would have to be designed and implemented with care. There is a good reason why historians tend to shy away from "what if?" scenarios, and similar reasons apply to scientific what-ifs, lest granting agencies are deluged with irresponsible submissions. Even those who think that Kuhn's account of normal science was too rigid (as I do) can agree with him that it would not be healthy for too many things to be called into serious question at once, lest mature sciences begin to resemble philosophy! On the plus side, however, scenario planning avoids (or at least mitigates) the Kuhnian objection that deviation from the established paradigm would undermine the enterprise of science. As-if thinking applied to the middle range need not directly threaten current scientific orthodoxy anymore than it upsets business orthodoxy, while still encouraging some in the expert community to retain an open future, one in which current orthodoxy will eventually seem badly dated.

Peter Drucker, whom I have quoted in Sect. 5, once discriminated four ways in which a company could improve its performance [17]:

(a) Abandon inefficient practices.
(b) Improve performance of retained practices through training.
(c) Improve via evolutionary innovation.
(d) Improve via innovation "to create the different tomorrow that makes obsolete and, to a large extent, replaces even the most successful products of today in any organization."

Item (d) is different from old-style planned obsolescence. It is the attitude of rapidly moving, highly innovative companies and of translational research, but it is much rarer in the sphere of basic research and absent in end-of-history thinking.

To be sure, basic scientific research is different in important ways from business enterprise, but Drucker's advice is surely relevant to basic scientific research. Implementing (d) would encourage a greater degree of transformational thinking if investigators and granting agencies could be persuaded to think of the various sciences as ongoing processes of technological design instead of as representations of the world that are already very close to expressing the final truth in nature's own language. After all, scientific models, theories, data sets, and, of course, experimental designs are all *designs*, created by human beings. Although they must survive the constraint of severe testing against nature,[21] and although some investigators will find it helpful to attribute truth, or approximate truth or validity,

[21] Of course, successful technological innovation must also function well enough to enjoy a viable market share, although, thanks to timing and marketing advantages, the objectively superior technology does not always win. Is science different in this respect?

to some of them, I am convinced that it would benefit the research enterprise to adopt a displacement model of the future. For, again, it seems most unlikely that the future of even mature sciences will be a flat extension of the present. We should not allow that perception to continue as the default view just because HA is messier to implement than CA.

12 Some Questions—and Two Final Thoughts

The questions express a double irony. Philosophy has traditionally been a normative discipline, one presumably, therefore, well equipped to engage in policy discussions. And at least since Plato's Meno paradox, the problem of the growth of knowledge has been a central philosophical issue. By contrast, the other science studies disciplines have claimed to be descriptive, largely shunning normative pretensions. Moreover, they have shied away from "context of discovery," since, to many science studies practitioners, 'discovery' is a philosopher's code word for strong realism. So why is it that these other science studies disciplines today have more purchase with policy makers than philosophy of science does? This is a change from the past. For example, following World War II, the U.S. National Science Foundation was theorized in the pages of the journal *Philosophy of Science*.[22] "Science studies" in its current form did not yet exist.

I am far from the first person to raise this question. The answer surely lies both in the academic success of science and technology studies (STS) and in philosophers' refusal to join—in other words, what they have done to themselves. On the success side, much work in STS does fall under what we may call "frontier epistemology" or "context of ongoing inquiry" as opposed to confirmation theory. But here I am more interested in the philosophical reasons. Because of the traditional self-understandings of philosophers of science and of the particular way in which the internal dynamics of the field of philosophy (founded by empiricists who were not themselves empirical, given their love for abstract, a priori calculi) of science have played out, philosophy of science narrowed itself to the point of simply "giving away" both areas—policy and frontier innovation—to the other science studies disciplines. There is a connection between the two giveaways. Arguably, the lack of attention to context of ongoing inquiry and to history, as well as adherence to an over-intellectualized distinction of pure versus applied science, produced an overreaction against the premature normativity of the logical positivists.

The next question is, What can philosophers of science do to reclaim some of this territory as for themselves? It is not a question of grabbing back something that has been taken from them. Insofar as the other science studies have succeeded in these areas, more power to them! Rather, it is a matter of philosophers developing their own stances in these neglected areas, not in order simply to gain social visibility but

[22] Thanks to Don Howard for reminding me of this.

because they (we), too, have an obligation to address problems of science, technology, and society. Writing articles on traditional topics for the journals is not enough. Fortunately, this question is now on the way to being answered by a generation of philosophers of science who are no longer so shy about making normative claims in the wake of the historical debunking of much of earlier philosophy of science. There are several recent books and articles that address policy issues.

I conclude with my two final thoughts. First, most philosophers of science have treated scientific progress primarily as increase in knowledge-that rather than in knowledge-how, as if science, like Christian theology, is more a matter of belief than of skilled modes of inquiry as an ongoing process. I do not wish to deny the importance and intellectual excitement of new propositional claims about what the universe is like, but we surely need more attention to expertise, including that involved in judicious HA. After all, at the furthest frontiers of research what most demarcates the scientific expert from the layperson, including such critics of science as creationists, is not deep propositional knowledge about the domain; for no one knows much about the domain, propositionally, at that point. Rather, it is in knowing *how* to conduct research, including HA of the various proposals for fruitful inquiry (see [51]). To be fair, such expertise does rely on a good deal of background knowledge, much of it propositional. That is not the same, however, as saying that this background knowledge is knowledge of the final theoretical *truth* about the world. In this contribution I have tried to remain neutral on that question.

Second, once we take seriously the idea that anything contributing to the growth of knowledge, including its *rate* of growth, is epistemically relevant, our notion of what is epistemically relevant must broaden considerably beyond the "context of justification" conception of traditional confirmation theory. And here the messy subject of HA becomes our guide. This means breaching the old internal-external distinction in important respects, for many of the factors clearly relevant to the rate of research progress are what philosophers traditionally have considered "external" and hence philosophically irrelevant and uninteresting. Philosophy should widen its horizons in this respect, as the other science studies already have done. Commitment to the old internal/external distinction in this context is a detriment to research progress as well as to research policy. And, once again, philosophers could take some clues from rapidly innovative sectors of the business world. While many philosophers regard such suggestions as radical, it is not as crazy as it probably sounds to them, for we can continue to distinguish different strands of epistemic relevance. This should include those components of HA that key on technical possibilities. A technical consideration within a domain of expertise remains distinct from a financial consideration or a political consideration. Since public science policy must be based on a wider sort of cost-benefit analysis than that usually considered by philosophers, there is an important role for HA to play, in the broad manner in which I conceive HA.

Acknowledgements Thanks to Emiliano Ippoliti for organizing the workshop and also to Maria Teresa Cipollone for stimulating questions. Unfortunately, given the length of the present paper, I have had to postpone my published answer to her. I have benefitted from recent correspondence with Britt Holbrook and from sessions on policy at the February 2013 American

Association for the Advancement of Science (AAAS) meetings in Boston. My largest debt is to the MIRRORS project at the University of Catania, Italy, which first got me, hesitantly, into the policy business. Thanks there especially to Franco Coniglione, Salvo Vasta, and Enrico Viola (see [10, 70]). Enrico and I continue to collaborate on these issues. And thanks to Roberto Poli for calling my attention to scenario planning some years ago. For helpful discussion of realism issues, i am indebted to my students, especially to Devin Bray.

References

1. American Academy of Arts and Sciences: Arise 2: Unleashing America's Research & Innovation Enterprise. Am. Academy of Arts and Sciences, Cambridge (2013)
2. Beinhocker, E.: The Origin of Wealth: The Radical Remaking of Economics and What It Means for Business and Society. Harvard Business School Press, Cambridge (2006)
3. Benkler, Y.: The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press, New Haven (2006)
4. Boyd, R.: On the current status of scientific realism. In: Boyd, R., Gasper, P., Trout, J.D. (eds.) The Philosophy of Science. MIT Press, Cambridge (1991)
5. Campbell, D.T.: Evolutionary epistemology. In: Schilpp, P.A. (ed.) The Philosophy of Karl R. Popper, pp. 413–463. Open Court, La Salle (1974)
6. Cellucci, C.: Perchè Ancora la Filosofia?. Laterza, Rome (2008)
7. Cellucci, C.: Philosophy of mathematics: making a fresh start. Stud. Hist. Philos. Sci. **44**, 32–42 (2013)
8. Christensen, C.: The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. Harvard Business School Press, Cambridge (1997)
9. Cohen, I.B.: Revolution in Science. Harvard University Press, Cambridge (1985)
10. Coniglione, F. (ed.): Nello specchio della scienza. Mondadori, Torino (2009)
11. De Langhe, R.: The division of labour in science: the tradeoff between specialisation and diversity. J. Econ. Methodol. **17**(1), 37–51 (2010)
12. De Langhe, R.: A Unified Approach to the Organization of Cognitive Labor. Philosophy of Science, forthcoming
13. Dennett, D.C.: Darwin's Dangerous Idea: Evolution and the Meanings of Life. Simon & Schuster, New York (1995)
14. Devitt, M.: Are unconceived alternatives a problem for scientific realism? (review of Stanford, P.K., exceeding our grasp). J. Gen. Philos. Sci. **42**, 285–293 (2011)
15. Dubois, L.: DARPA's approach to innovation and its reflection in industry. In: Reducing the Time from Basic research to Innovation in the Chemical Sciences: A Workshop Report to the Chemical Sciences Roundtable, National Research Council, pp. 37–48. National Academies Press, Washington, DC (2003) www.nap.edu/catalog/10676.html
16. Drucker, P.: The Age of Discontinuity: Guidelines to Our Changing Society. Harper & Row, New York (1968)
17. Drucker, P.: Management's New Paradigms. Forbes (Oct 5), pp. 152–177 (1998)
18. Etzkowitz, H.: The Triple Helix: University-Industry-Government Innovation in Action. Routledge, London (2008)
19. Feigenbaum, E., Buchanan, B., Lederberg, J.: On generality and problem solving: a case study using the DENDRAL program. Mach. Intell. **7**, 165–190 (1971)
20. 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, pp. 305–342. Lawrence Erlbaum Associates, Mahwah (2007)
21. Gillies, D.: Artificial Intelligence and Scientific Method. Oxford University Press, Oxford (1996)

22. Godfrey-Smith, P.: Is it a revolution? Biol. Philos. **22**(3), 429–437 (2007)
23. Henderson, L., Goodman, N., Tenenbaum, J., Woodward, J.: The structure and dynamics of scientific theories: a hierarchical bayesian perspective. Philos. Sci. **77**, 172–200 (2010)
24. Hesse, M.B.: Truth and the growth of knowledge. In: Suppe, F., Asquith, P.D. (eds.) PSA 1976, vol. 2, pp. 261–280. Philosophy of Science Association, East Lansing (1976)
25. Ippoliti, E.: Il Vero e il Plausibile. Lulu, Morrisville (2008)
26. Ippoliti, E.: Inferenze Ampliative: Visualizzazione, Analogia, e Rappresentazioni Multiple, 2nd edn. Lulu, Morrisville (2012)
27. Jablonka, E., Lamb, M.: Evolution in Four Dimensions. MIT Press, Cambridge (2005)
28. Kay, L.: The Molecular vision of Life: Caltech, the Rockefeller Foundation, and the Rise of the New Biology. Oxford University Press, New York (1993)
29. Kuhn, T.S.: The Structure of Scientific Revolutions. University of Chicago Press, Chicago (1962), 2nd edn. with postscript (1970)
30. Kuhn, T.S.: The Essential Tension: Selected Studies in Scientific Tradition and Change. University of Chicago Press, Chicago (1977)
31. Kuhn, T.S.: Black-Body Theory and the Quantum Discontinuity, 1894–1914. Clarendon Press, Oxford (1978)
32. Latour, B.: Science in Action: How to Follow Scientists and Engineers through Society. Oxford University Press, Oxford (1987)
33. Laudan, L.: Progress and Its Problems. University of California Press, Berkeley (1977)
34. Levins, R.: The Strategy of Model Building in Population Biology. Am. Sci. **54**, 421–431 (1966)
35. Lindgren, M., Bandhold, H.: Scenario Planning: The Link between Future and Strategy. Palgrave Macmillan, Houndmills, Hampshire (2003)
36. Magnani, L., Nersessian, N., Thagard, P. (eds.): Model-Based Reasoning in Scientific Discovery. Kluwer, Dordrecht (1999)
37. Margolis, H.: Patterns, Thinking, and Cognition. University of Chicago Press, Chicago (1987)
38. Margolis, H.: Paradigms and Barriers: How Habits of Mind Govern Scientific Beliefs. University of Chicago Press, Chicago (1993)
39. McMullin, E.: The fertility of theory and the unit for appraisal in science. In: Cohen, R.S., et al. (eds.) Essays in Memory of Imre Lakatos, pp. 395–432. Reidel, Dordrecht (1976)
40. Meheus, J. (ed.): Inconsistency in Science. Kluwer, Dordrecht (2002)
41. Merton, R.K.: Social Theory and Social Structure. Free Press, New York (1968)
42. Newell, A., Simon, H.A.: Human Problem Solving. Prentice-Hall, Englewood Cliffs (1972)
43. Nickles, T.: Heuristic appraisal: a proposal. Soc. Epistemology **3**, 175–188 (1989)
44. Nickles, T.: Evolutionary models of innovation and the meno problem. In: Shavinina, L. (ed.) The International Handbook on Innovation, pp. 54–78. Pergamon, Oxford (2003)
45. 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)
46. Nickles, T.: The strange story of scientific method. In: Meheus, J., Nickles, T. (eds.) Models of Discovery and Creativity, pp. 167–207. Springer, Dordrecht (2009)
47. Nickles, T.: Vantaggio Competitivo e Futuro del Capitalismo: Uno Squardo al Contesto. In: Coniglione, F. (ed.) Nello Specchio della Scienza. Mondadori, Milano (2009)
48. Nickles, T.: Life at the frontier: the relevance of heuristic appraisal to policy. Axiomathes **19**, 441–441 (2009)
49. Nickles, T.: Entrepreneurship and Frontier Theory of inquiry. In: Viola, E. (ed.) Epistemologies and the Knowledge Society: New and Old Challenges for 21st-century Europe, pp. 31–86. Bonanno, Rome (2010)
50. Nickles, T.: Scientific revolutions. In: Zalta, E. (ed.) Stanford Encyclopedia of Philosophy Winter (2013 edition). http://plato.stanford.edu/archives/win2013/entries/scientific-revolutions/
51. Nickles, T.: The problem of demarcation. In: Pigliucci, M., Boudry, M. (eds.) Philosophy of Pseudoscience: Reconsidering the Demarcation Problem, pp. 101–120. University of Chicago Press, Chicago (2013)

52. Nietzsche, F.: On the uses and disadvantages of history for life. In Hollingdale, R.J. (trans.) Untimely Meditations, pp. 59–123. Cambridge University Press, Cambridge (1983)
53. Plotkin, H.: Darwin Machines and the Nature of Knowledge. Harvard University Press, Cambridge (1997)
54. Popper, K.R.: The Poverty of Historicism, 2nd edn. Routledge, London (1961)
55. Popper, K.R.: Objective Knowledge: An Evolutionary Approach. Clarendon Press, Oxford (1972)
56. Post, H.R.: Correspondence, invariance, and heuristics. In praise of conservative induction. Stud. Hist. Philos. Sci. **2**, 213–255 (1971)
57. Psillos, S.: Scientific Realism: How Science Tracks Truth. Routledge, London (1999)
58. Ralston, B., Wilson, I.: The Scenario Planning Handbook. Thomson-South-Western, Mason (2006)
59. Rescher, N.: Scientific Progress and the 'Limits of Growth'. Chapter 8 of Rescher's Unpopular Essays on Technological Progress, pp. 93–104. University of Pittsburgh Press, Pittsburgh (1980)
60. Rescher, N.: Nature and Understanding: The Metaphysics and Method of Science. Clarendon Press, Oxford (2000)
61. Rheinberger, H.-J.: Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube. Stanford University Press, Stanford (1997)
62. Saxenian, A.: Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Harvard University Press, Cambridge (1996)
63. Schumpeter, J.: Capitalism, Socialism, and Democracy. Harper, New York (1942)
64. Šešelja, D., Christian Straßer, C.: Kuhn and coherentist epistemology. Stud. Hist. Philos. Sci. Part A **40**(3), 322–327 (2009)
65. Šešelja, D., Straßer, C.: Kuhn and the question of pursuit worthiness. Topoi **32**(1), 1–19 (2012)
66. Shapere, D.: Notes Toward a Post-Positivistic Interpretation of Science, Part II. Chapter 17 of Shapere's Reason and the Search for Knowledge, pp. 352–382. Reidel, Dordrecht (1984)
67. Simon, H.A.: Discussion: the meno paradox. Philos. Sci. **43**, 147–151 (1976). Reprinted in Simon's Models of Discovery, pp. 338–341. Reidel, Dordrecht (1977)
68. Stanford, P.K.: Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives. Oxford University Press, New York (2006)
69. van Fraassen, B.C.: The Empirical Stance. Yale University Press, New Haven (2002)
70. Viola, E. (ed.): Epistemologies and the Knowledge Society: New and Old Challenges for 21st-century Europe. Bonanno, Rome (2010)
71. von Hippel, E.: Democratizing Innovation. MIT Press, Cambridge (2005)
72. Wagner, H.: The open corporation. Calif. Manag. Rev. **33**(4), 46–60 (1991)
73. Warsh, D.: Knowledge and the Wealth of Nations: A Story of Economic Discovery. Norton, New York (2006)
74. Wolpert, D.: The lack of a priori distinctions between learning algorithms. Neural Comput. **8**, 1341–1390 (1996)
75. Wolpert, D., Macready, W.: No Free Lunch Theorems for Search. Technical Report SFI-TR-95-02-010. Santa Fe Institute, Santa Fe, NM (1995)
76. Wolpert, D., Macready, W.: No free lunch theorems for optimization. IEEE Trans. Evol. Comput. **1**, 67 (1997)

Why Do Scientific Revolutions Begin?

Donald Gillies

Abstract This paper is concerned with the problem of why scientific revolutions begin. It considers first Kuhn's view that a revolution is started by a build-up of anomalies in the old paradigm. This view is criticized on historical grounds by considering the examples of the Einsteinian revolution and the Copernican revolution. It is argued that there was no significant build-up of anomalies in the old paradigm just before the beginning of these revolutions. An alternative view is then put forward that the start of a revolution has to be explained in terms of technology and practical problems (or tech for short). There are two patterns: (i) *tech first* in which technological advances lead to new discoveries and these lead to the onset of the revolution, and (ii) *tech last* in which the need to solve an urgent practical problem produces a challenge to the old paradigm. If this challenge is successful, the new paradigm leads to a solution of the practical problem and so to technological advance. The tech first pattern is illustrated by the example of the chemical revolution, and the tech last pattern by the example of the development of the germ theory of disease. It is then argued that scientific revolutions can exhibit a combination of tech first and tech last, and this is illustrated by the Copernican revolution. In the final section of the paper, it is shown that the 'tech first/tech last' theory explains why the Copernican revolution occurred in Europe in the 16th and 17th centuries, and not in the ancient Greek world (with Aristarchus), or in China in the 16th and 17th centuries.

1 Introduction. The Problem

According to Kuhn's model, in most branches of science for most of the time, the research scientists all accept the dominant paradigm of the field, and carry out normal science within the framework of that paradigm. Occasionally, however, a few

D. Gillies (✉)
University College London, London, UK
e-mail: donald.gillies@ucl.ac.uk

E. Ippoliti (ed.), *Heuristic Reasoning*, Studies in Applied Philosophy, Epistemology and Rational Ethics 16, DOI 10.1007/978-3-319-09159-4_5

of these research scientists challenge the dominant paradigm and try to develop a new one. If they are successful, we have a scientific revolution. The question I want to raise in this paper is why, in the midst of the usual normal science, do these occasional challenges to the dominant paradigm arise? Kuhn himself suggests an answer this question. Section 2 will state and criticize this proposed solution to the problem.

2 Critique of Kuhn's 'Build-up of Anomalies' View

Kuhn's proposed solution uses his concept of anomaly. Regarding the concept of anomaly, Kuhn writes [17, pp. 52–53]:

> Discovery commences with the awareness of anomaly, i.e. with the recognition that nature has somehow violated the paradigm-induced expectations that govern normal science.

Presumably what Kuhn has in mind is something like this. From the assumptions of the paradigm, together with what seem to be plausible auxiliary conditions, a result is deduced which is contradicted by observation. This contradiction constitutes an anomaly.

Using this concept, Kuhn writes about the beginning of scientific revolutions as follows [17, p. 6]:

> … normal science repeatedly goes astray. And when it does – when, that is, the profession can no longer evade anomalies that subvert the existing tradition of scientific practice – then begin the extraordinary … episodes … known in this essay as scientific revolutions.

I will characterize this idea of Kuhn's as the 'build-up of anomalies' view. If an increasing number of anomalies occur in the dominant paradigm, this will lead to the paradigm being questioned, and a departure from normal science. This view, which has a slightly Popperian flavour, seems very plausible and in accordance with common sense. Unfortunately, however, it does not agree well with historical facts. There may be periods when the dominant paradigm is beset with many anomalies, and yet normal science continues unchallenged. Conversely, scientific revolutions can sometimes begin when there are only a few, and rather minor, anomalies in the dominant paradigm.

In order to argue for these claims, I will consider one of the classic examples of normal science, namely research in astronomy and mechanics from the time of the general acceptance of Newton's theory (c. 1720) to the beginning of the Einsteinian revolution (c. 1905). In fact there were in the Newtonian normal science of this period a succession of anomalies which often remained unresolved for quite long periods of time. In 1746, for example,[1] Clairaut found that the progress of the Moon's apogee is twice what it should be according to Newton's theory. It turned out that Clairaut had exaggerated the size of this anomaly owing to a

[1] This example is discussed in Lakatos [18, p. 219].

mathematical mistake, but there still remained the small anomaly of a 'secular acceleration'. This was only successfully resolved within Newtonian theory by Laplace in 1787. The next anomaly to come to light concerned the planet Uranus which was discovered in 1781. The first to compute its orbit was Lexell, and he noticed that it had irregularities. In 1821 Bouvard made predictions using Newtonian theory of the future positions of Uranus, but subsequent observations revealed substantial deviations from Bouvard's theoretical values. In 1846 this anomaly was triumphantly resolved by Adams and Leverrier. They explained the irregularities in the orbit of Uranus by postulating a hitherto unknown planet, and then used Newtonian theory to calculate where that planet should be. The new planet (Neptune) was duly observed on 23 September 1846 only 52′ away from the predicted position. Leverrier, having resolved one anomaly in Newtonian normal science, went on to discover another. In 1859 he showed that the rate of precession of the perihelion of Mercury differed from that predicted by Newtonian theory by 38″ per century. Later his estimate of 38″ was changed to 43″. Leverrier tried to explain this anomaly in the same way that he dealt with the irregularities in the orbit of Uranus. He postulated a hitherto unknown planet nearer the Sun than Mercury. This hypothetical planet was even given the name 'Vulcan', but no such planet was ever discovered. In fact the anomaly of the rate of precession of the perihelion of Mercury was never resolved within the Newtonian paradigm. However, as the anomaly was a tiny one, and as similar anomalies had been successfully resolved on earlier occasions, it is unlikely that this anomaly reduced confidence in the Newtonian paradigm to any significant extent.

Our analysis of this example shows that anomalies can frequently arise in normal science and can in some cases remain unresolved for long periods without giving rise to a scientific revolution or significantly reducing confidence in the dominant paradigm. However, it is still possible that just before a scientific revolution, there is a build-up of a large number of anomalies, and it is this build-up, which triggers the revolution. Let us next examine this hypothesis as applied to the Einsteinian revolution.

When the Einsteinian revolution began about 1905, the dominant paradigm was no longer just Newtonian theory, but a combination of Newtonian theory with Maxwell's electrodynamics. The two theories seemed to fit well together. Maxwellian theory postulated the existence of an ether, and regarded electromagnetic radiation as waves in this ether. Now the ether could provide a basis for the absolute space which Newton had postulated. In the Einsteinian revolution, however, this paradigm was replaced by one in which the existence of both the ether and absolute space was denied, and in which Newtonian mechanics and gravitational theory were replaced by the special and general theories of relativity. On the build-up of anomalies theory, we would expect that around 1900 there would be a large number of anomalies in the Newton-Maxwell paradigm. Was this in fact the case?

As far as Newtonian theory is concerned the only anomaly was that concerned with the perihelion of Mercury which had been known since 1859, and which no one considered to be very serious. What about the ether? At the International Congress of Physics, held in Paris in 1900, Lord Kelvin gave an address in which he considered ether theory. He remarked that "the only cloud in the clear sky of the theory was the null result of the Michelson-Morley experiment."[2] So Lord Kelvin, one of the leading physicists of the time only recognised one anomaly in the ether theory.

However, it could be claimed that Lord Kelvin was wrong to consider this to be an anomaly because it had been successfully explained in terms of the dominant paradigm. In his 1892 [24] and 1895 [25], Lorentz had explained the null result of the Michelson-Morley experiment of 1887 using the contraction hypothesis. As this hypothesis had been put forward independently by Fitzgerald, it became known as Lorentz-Fitzgerald Contraction (or LFC). Now for many years LFC was dismissed as a purely ad hoc hypothesis, which did not satisfactorily explain the result of the Michelson-Morley experiment. This point of view is to be found in Popper [31, Sect. 20, p. 83]. However, Grünbaum [11] argued that the Lorentz-Fitzgerald Contraction was not an ad hoc hypothesis. This criticism was accepted by Popper [32] and Lakatos [19, p. 75, Footnote 5]. However, Holton [13] continued to maintain that the LFC was an ad hoc hypothesis, though he used ad hoc in a different sense from Popper. Holton's view was in its turn criticized very convincingly by Zahar [34, pp. 5–10, 62–66]. Zahar showed that the LFC was deduced by Lorentz from another deeper hypothesis—his Molecular Forces Hypothesis, which Lorentz had introduced for reasons which had nothing to do with the Michelson-Morley experiment. Zahar concluded from this that the LFC was not ad hoc in Holton's sense of the term. However, if the LFC was not ad hoc, it provided a satisfactory resolution of the anomaly created by the null result of the Michelson-Morley experiment. It follows that in 1900 there was only one anomaly in the dominant Newton-ether paradigm (P_1), namely the tiny anomaly of the rate of precession of the perihelion of Mercury, an anomaly, which had been known since 1859. In effect there was no build-up of anomalies in 1900, only 5 years before the beginning of the Einsteinian revolution.

Let us take as our second example the Copernican revolution. This is usually regarded as beginning with the publication of Copernicus' *De Revolutionibus Orbium Caelestium* in 1543. What was the state of the dominant paradigm in astronomy at that time. Kuhn answers as follows [17, p. 67]:

> On this point historical evidence is entirely unequivocal. The state of Ptolemaic astronomy was a scandal before Copernicus' announcement.

I cannot agree with Kuhn here. On the contrary, at that time there seem to have been hardly any anomalies in the dominant Ptolemaic paradigm. With its apparatus of cycles and epicycles, Ptolemaic astronomy could explain nearly all the observed phenomena, and also predict the movements of the planets fairly accurately. Of course the accuracy was much less than would be achieved later, but at the time, Ptolemaic astronomy was undoubtedly the most accurate of all the existing sciences.

The best evidence for the absence of anomalies in Ptolemaic astronomy in 1543 is provided by the text of Copernicus' *De Revolutionibus*. In the Preface and Book 1,

[2] Miller [28, p. 618]. This reference comes from Lakatos [19, p. 72, Footnote 6].

Copernicus set out his main arguments against Ptolemaic astronomy. If there had been any relevant anomalies in that system, he would surely have mentioned them. Yet he does not do so. In the Preface he complains that Ptolemaic astronomy is not sufficiently accurate, but here he is surely on weak grounds since the accuracy was very good for the time. He also complains about the complexity and confusion of the mathematical methods used by Ptolemaic astronomers. He particularly objected to the use of equants, writing (p. 150):

> Those … who have devised eccentric systems … have yet made many admissions which seem to violate the first principle of uniformity in motion.

Copernicus himself did avoid the use of equants in his work, but the mathematical complexity of his system was in fact just as great as that of Ptolemy.

In Book 1 itself, Copernicus mentions a number of astronomical facts, which are more simply explained in his system. One such fact is that Mercury and Venus, unlike the other planets, always stay close to the Sun [4, p. 167]. Another fact is that the outer planets (Mars, Jupiter, and Saturn) are always brightest and hence nearest to the Earth when in opposition to the Sun, and most distant when in conjunction with the Sun [4, pp. 167–168]. Copernicus can also easily explain the phenomenon of planetary retrogression, and also a number of facts about retrogressions, which he mentions in the following passage [4, pp. 169–170]:

> For here we may observe why the progression and retrogression appear greater for Jupiter than Saturn, and less than for Mars, but again greater for Venus than for Mercury; and why such oscillation appears more frequently in Saturn than in Jupiter, but less frequently in Mars and Venus than in Mercury; … All these phenomena proceed from the same cause, namely Earth's motion.

Copernicus is quite right here. The facts that he mentions can indeed be explained simply on his heliocentric hypothesis. However, what I would like to stress, is that he does not say of any these facts that it cannot be explained on the Ptolemaic system, i.e. that it is an anomaly for that system. Nor would he have been correct to do so. Ptolemaic astronomers were able to explain all these facts, though in a more complicated fashion, by manipulating cycles and epicycles.

These then are my historical arguments against the 'build-up of anomalies' theory of why scientific revolutions begin. I now turn to giving my alternative theory. I do not think the beginning of a scientific revolution can be explained entirely in terms of the situation internal to that science. We need to bring into the picture some things, which are external, namely technology and practical problems. For convenience I will use 'tech' as an abbreviation for technology and practical problems. There are, I claim, two different ways in which tech can give rise to a scientific revolution. Advances in technology can precede a scientific revolution. Usually the relevant effect of the new technology is to enable better instruments for scientific use to be constructed. This enables new observations and experiments to be carried out, resulting in a number of significant discoveries, which involve new objects and new processes. These discoveries are what give rise to the scientific revolution. I call this pattern: 'tech first', because the advances come before the scientific revolution, and act as its efficient cause. By contrast there is also a 'tech last' pattern, which takes the following form. At a certain stage of development, there may be some urgent practical problems, which cannot be easily solved within the existing scientific paradigm. Some scientists may be stimulated by this situation to try to solve these problems by changing the paradigm. This explains the beginning of the scientific revolution. If the new paradigm does indeed produce solutions to the urgent practical problems, then the scientific revolution will be successful. I call this pattern 'tech last', because the advances in tech occur after the scientific revolution has begun, and, as a consequence of the scientific revolution. Tech is here the final cause rather than the efficient cause of the revolution.

Interestingly Kuhn himself came to have doubts about his 'build-up of anomalies' theory.[3] As Hoyningen-Huene points out in his [15, pp. 232–233]:

> The thesis that all revolutions in theory are indicated by crises, triggered in turn by the appearance of significant anomalies in the relevant field, has, subsequent to being met with criticism, been somewhat weakened. Kuhn remains, as before, convinced that crises are usually the prelude to revolution, but he acknowledges that revolutions might also, albeit rarely, get started in other ways.

In the Postscript, written in 1969, to the second edition of *The Structure of Revolutions*, Kuhn writes [17, p. 181]:

> A number of critics[4] have doubted whether crisis, the common awareness that something has gone wrong, precedes revolutions so invariably as I implied in my original text. Nothing important to my argument depends, however, on crises' being an absolute prerequisite to revolutions; they need only be the usual prelude, supplying, that is, a self-correcting mechanism which ensures that the rigidity of normal science will not forever go unchallenged. Revolutions may also be induced in other ways, though I think they seldom are. … crises need not be generated by the work of the community that experiences them and that sometimes undergoes revolution as a result. New instruments like the electron microscope or new laws like Maxwell's may develop in one specialty and their assimilation create crisis in another.

Here Kuhn mentions the role of instruments, and, he also refers to technology in another passage. He says [17, p. x]:

> I have said nothing about the role of technological advance or of external social, economic, and intellectual conditions in the development of the sciences. One need, however, look no further than Copernicus and the calendar to discover that external conditions may help to transform a mere anomaly into a source of acute crisis.

To some extent then this paper takes up these hints of Kuhn's concerning the role of technology and tries to develop them.

In Sect. 3, I will describe the 'tech first' pattern in more detail, and use it to explain the beginning of the chemical revolution. In Sect. 4, I will treat the 'tech last' pattern in a similar fashion, using as an example a revolution in medicine.

---

[3] I am very grateful to Thomas Sturm for pointing this out to me, and also supplying the references to Hoyningen-Huene and Kuhn.

[4] Unfortunately Kuhn does not say who these critics were.

3 Tech First, and the Beginning of the Chemical Revolution

To illustrate the concept of 'tech first' I will begin by giving a simple and very striking example which was not actually the beginning of a scientific revolution, though it did give a boost to a scientific revolution already under way. This example is Galileo's telescopic discoveries.

Galileo seems to have heard a report of a telescope made in Belgium in June 1609, when he was teaching at the University of Padua in the Venetian republic. He succeeded in making a telescope for himself in July and August of 1609. Then in March 1610 he published the first report of his astronomical observations using his telescope in a pamphlet called *Sidereus Nuncius* (The Starry Messenger).

The discoveries, which Galileo made in such a short space of time with his new instrument, were truly remarkable. First of all he found that there were mountains on the moon, and even gave a quite accurate estimate of the height of the tallest of them (about 4 miles). Secondly he was able to observe thousands of previously unknown stars. For example, he says [8, p. 47]:

> Hence to the three stars in the Belt of Orion and the six in the Sword which were previously known, I have added eighty adjacent stars …

Thirdly, and perhaps most strikingly of all, he discovered that Jupiter had 4 moons.

This example is a perfect illustration of what I call the 'tech first' pattern. Technological developments lead to new instruments, and, with the help of these, a number of striking new discoveries are made. I claim that this 'tech first' pattern is, in some cases, what stimulates the beginning of a scientific revolution. It might be said then that in such cases, I am replacing a build up of anomalies theory with a build up of new discoveries theory. This is quite correct, and it is therefore important to explain how anomalies differ from new discoveries in relation to the dominant paradigm in the branch of science in question.

As we saw from our analysis of Newtonian normal science in the period c. 1720–c. 1900, anomalies are concerned with objects, which are standardly dealt with in the paradigm. In the examples we considered these were the Moon, planets such as Uranus and Mercury, and, in the Michelson-Morley experiment, the ether, which was a basic object in the extended Newton-Maxwell version of the paradigm. Characteristically an anomaly arises because the paradigm predicts some result regarding the object, which is contradicted by observation, so that some adjustment becomes necessary. In the case of new discoveries, however, the objects may be quite new, and may never have been considered within the old paradigm. Moreover these new objects may behave in new and unfamiliar ways. How such new objects can be handled within the dominant paradigm thus becomes a much more problematic matter. This is well illustrated by the example of Galileo's telescopic discoveries.

The mountains on the Moon definitely contradicted one claim of the Aristotelian-Ptolemaic paradigm, namely that the heavenly bodies were perfect spheres. However, it is by no means clear how important this is for the rest of the paradigm. Perhaps the general Ptolemaic system could be maintained while abandoning the doctrine of the perfection of heavenly bodies. Then again consider the thousands of newly discovered stars. Could they be fitted into the old celestial sphere rotating round the Earth once a day? It doesn't seem impossible, but it is not so plausible either. Conversely, the fact that stars were not magnified in diameter by the telescope tended to support the view that they were very far away, an assumption which the Copernicans needed to explain the absence of stellar parallax.[5] Similar considerations apply to the moons of Jupiter, whose existence is much more naturally explained in the Copernican paradigm than the Ptolemaic. Moreover new discoveries of such a dramatic nature are bound to shift the mental attitudes of at least some researchers. The old paradigm, after all, was developed in complete ignorance of the new objects and phenomena. This is bound to suggest to some researchers (the revolutionaries) that a new paradigm is needed to deal with the new entities. Of course not all researchers will reason in this way. Other researchers (the conservatives) will try to explain the new discoveries in terms of the old paradigm. In this way a build up of new discoveries creates the conditions for the beginning of a scientific revolution. I will now try to illustrate this in the case of the chemical revolution.

The phlogiston theory was designed to explain two processes, which were regarded as essentially the same, namely combustion and calcination. An example of combustion would be the burning of charcoal. Calcination consisted of the conversion of a metal to its calx—for example, the conversion of iron into rust. These processes were explained by postulating that charcoal and metals were rich in an inflammable substance (phlogiston), which was expelled in the process of combustion or calcination. From the start, there was a problem about weight change in the phlogiston theory. When charcoal was burnt, there remained ashes, which weighed less than the charcoal. This was consistent with the assumption that a substance (phlogiston) had been expelled. However, when a metal was converted to its calx, the calx weighed more than the original metal. Supporters of the phlogiston theory explained away this anomaly in various ways. For example, Boyle and Boerhaave explained it, by supposing that in calcination fire particles enter the calx, and account for its increase in weight [21, p. 124]. The phlogiston theory could also explain the reverse of calcination, that is the conversion of calx into metal. For example if we heat the calx with charcoal, since charcoal is very rich in phlogiston, the phlogiston from the charcoal combines with the calx to give the metal.

What gave rise to the chemical revolution, which resulted in the overthrow of the phlogiston theory, was a build up of new discoveries, which were not in the field of combustion and calcination. These new discoveries were of new gases, whose chemical and physical properties were investigated. Although the word 'gas' had been invented by Van Helmont in the 17th century, it was not much used by chemists in the 18th century. What we now call a gas, they referred to as an 'air' or 'elastic fluid'.

The discoveries concerning new airs were in turn made possible by technological innovations in the devices used by chemists. Perhaps the most important of

[5] I owe this point to Andrew Gregory.

these was the invention of the pneumatic trough for collecting gases over water. This was due to Stephen Hales, a clergyman and amateur chemist, who published an account of it in his book *Vegetable Statics* of 1727.

Hales himself used his new apparatus with enthusiasm to obtain airs from a variety of substances, but he did not investigate the chemical properties of these airs.

Hales' pneumatic trough was developed by later scientists (Fig. 1). Figure 2 shows the form of the pneumatic trough used by Priestley for producing the gas, which we now call oxygen.

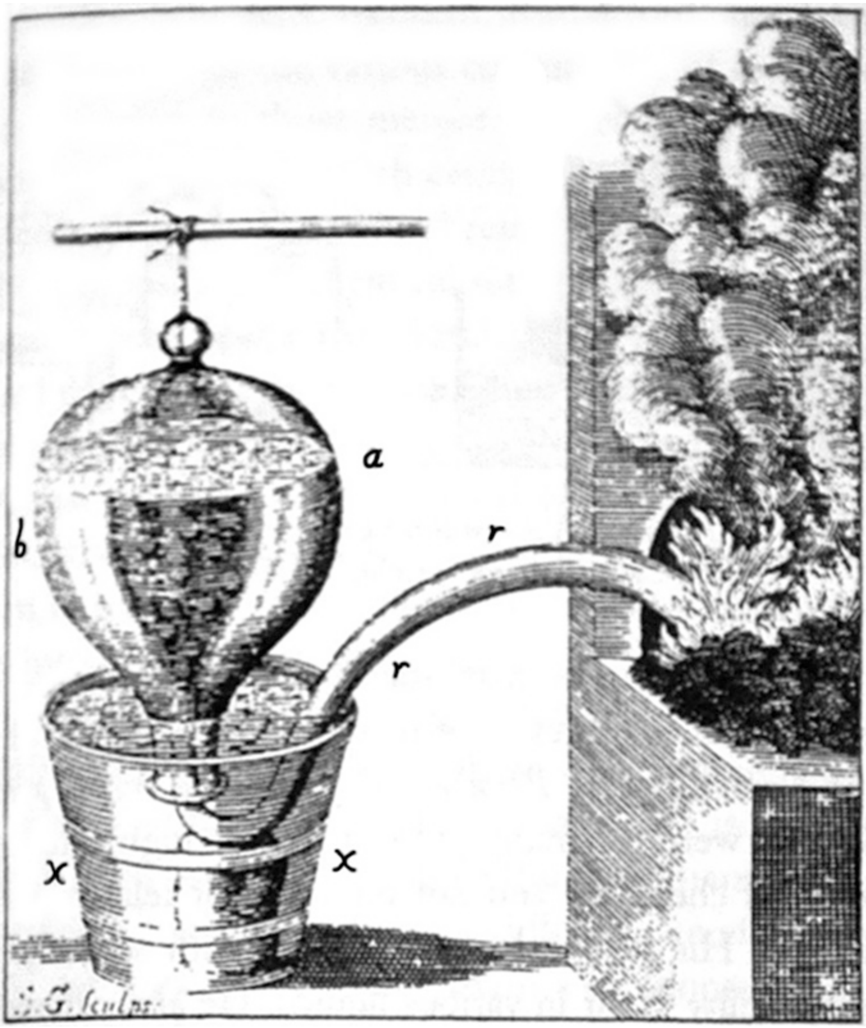


Fig. 1 Hales' pneumatic trough

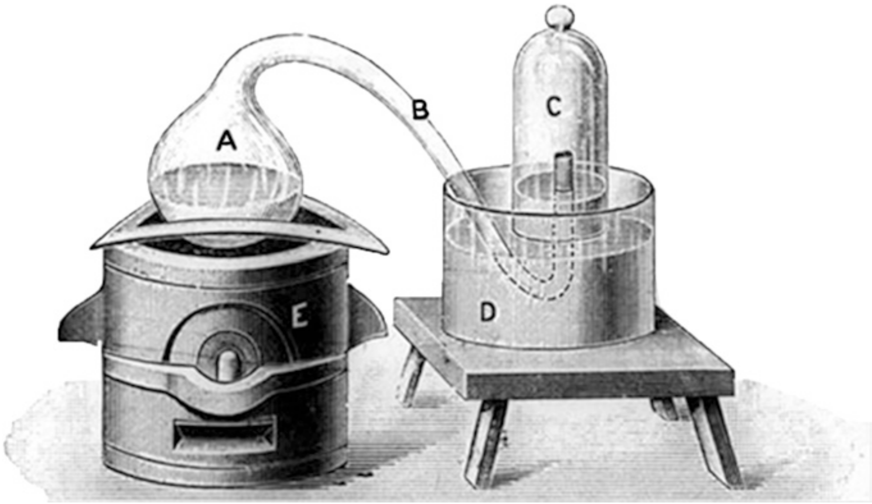


Fig. 2 Priestley's pneumatic trough

Both Cavendish and Priestley also modified the pneumatic trough for water soluble gases by collecting the gas over mercury rather than water.

Of course the pneumatic trough was only one of the pieces of apparatus, which the chemists of the time used for their investigations. Figure 3 shows the burning glass, which Priestley used for heating calx of mercury.

Improved balances were also important.

The next important advance after Hales was published by Joseph Black in 1756. He obtained a gas by heating lime, leaving a residue of quick-lime. This new gas Black called 'fixed air' (our carbon dioxide). However, Black also obtained the reverse reaction. If quick-lime is dissolved in water to get lime water, and then fixed air is bubbled through limewater, lime appears immediately as a milky precipitate. This suggests that the air has become fixed back into the lime—hence the name fixed air. The startling nature of this discovery is emphasised by Black's colleague John Robison who wrote in his introduction to the printed version of Black's lectures on chemistry, published posthumously in 1803:

> He had discovered that a cubic inch of marble consisted of about half its weight of pure lime and as much air as would fill a vessel holding six gallons. … What could be more singular than to find so subtile a substance as air existing in the form of a hard stone, and its presence accompanied by such a change in the properties of the stone? (quoted from [21, p. 134]).

Black also studied many of the properties of his fixed air. He knew that it extinguished flames, and that mice put in an atmosphere of fixed air died. He found that the density of fixed air was greater than that of common air, and tests with alkaline substances showed that it behaved like a weak acid. An unknown air could be identified as fixed air by these properties, and by the fact that it immediately created a milky precipitate when bubbled through limewater.

Fig. 3 Priestley's burning glass

Using these tests, Black was able to discover a number of different ways of producing fixed air (see [27, p. 37]). He found that fixed air was expired in respiration, and produced in alcoholic fermentation. He also found that fixed air could be produced by burning charcoal, or dissolving some mild alkalies in acid.

Black's pioneering work on fixed air was followed in the next twenty years by a series of discoveries of other new gases, and the study of their properties. In 1766 Cavendish published a study of what he called inflammable air (our hydrogen). In 1772 Rutherford, a student of Black, removed from common air all that could be eliminated by respiration and combustion. He recognized that what remained was a new air which he called 'mephitic air'. This is our nitrogen. In 1772 Priestley discovered 'nitrous air' and some other airs. The most interesting such air was one he obtained in August 1774 when he used a new burning lens to extract an air from *mercurius calcinatus* per se (our mercuric oxide). Priestley had to interrupt his work on this new air (our Oxygen) to accompany his patron Lord Shelburne on a continental tour. However, this had the advantage that he met Lavoisier in Paris in October 1774 and was able to tell him about the new air, which Lavoisier proceeded to investigate.

My thesis then is that this build up of discoveries concerning new gases and their properties gave rise to the chemical revolution. This thesis is supported by a most interesting document written by Lavoisier probably on 20 February 1773.[6] In this document Lavoisier describes his programme for research. He says (quoted from McKie [26, pp. 120–123]) that the aim of his programme is that of making a "long series of experiments … on the elastic fluid that is set free from substances, either by fermentation, or distillation or in every kind of chemical change, and also on the air absorbed in the combustion of a great many substances …". He mentions similar experiments by his predecessors, but regards them as inadequate: "However numerous may be the experiments of Messrs. Hales, Black, Magbride (Macbride—D.G.), Jacquin, Cranz, Prisley (Priestley—D.G.), and de Smeth, in this direction, nevertheless, they come far short of the number necessary for a complete body of doctrine. … I have been bound to look upon all that has been done before me as merely suggestive: I have proposed to repeat it all with new safeguards, in order to link our knowledge of the air that goes into combination or that is liberated from substances, with other acquired knowledge, and to form a theory." He also says prophetically: "the importance of the end in view prompted me to undertake all this work, which seemed to me destined to bring about a revolution in physics and chemistry." Lavoisier is remarkable in that he predicts that his research will bring about a revolution in chemistry, which was what indeed happened. He also thinks that it is the study of the new gases and their properties, which will give rise to this revolution. This is in accordance with the view of this paper.

The impact of the discovery of the new gases and their properties was different for different researchers. Lavoisier decided quite early on (perhaps as early as

[6] The document is actually dated February 20, 1772 by Lavoisier, but it occurs in a laboratory notebook for 1773. Most scholars think that 1773 is the correct date and that the date 1772 was a mistake by Lavoisier. For a discussion, see McKie [26, pp. 123–125].

November 1772) that the phlogiston theory would not explain the new results satisfactorily, and that the development of a new theory was needed. The leading English chemists (Cavendish and Priestley), on the other hand, tried to fit the new results into the framework of the old phlogiston theory. A conflict between the old paradigm and the emerging new paradigm developed over the next two decades and it ended with the victory of the new approach.[7]

Against this account, it might be objected that issues to do with combustion and calcination still remained central, and that, for example, Lavoisier's key experiments of 1772 on the burning of sulphur and phosphorus, rather than results about the new gases, were what convinced him of the falsity of the phlogiston theory. It is of course true that experiments on combustion and calcination remained central, but there was a close connection between these experiments and the new results concerning gases. Black had shown that limewater could be converted into lime through the fixing of his new air. This result favoured the view that combustion and calcination might involve the fixing of air, to produce an addition theory of combustion and calcination as opposed to the subtraction theory of phlogiston.

Earlier we remarked that the phlogiston theory explained the weight loss when charcoal was burned and turned to ashes. However the weight gain when a metal was turned into a calx was an anomaly. With an addition theory the situation was exactly reversed. The weight gain when a metal turned to calx could easily be explained, but the weight loss when charcoal is burned constituted an anomaly. However, some of Black's results concerning fixed air indicated a way in which this anomaly could be resolved. Black had shown that fixed air is produced when charcoal is burned. Perhaps when account is taken of this fixed air, there will be a weight gain rather than a weight loss. Lavoisier must have realised this early on, and he gives this explanation of the weight loss on the combustion of charcoal in his *Elements of Chemistry* [20, pp. 63–64]. Thus the new results on gases, not only suggested an addition theory, but also showed a way to resolve the anomaly in such a theory.

So to sum up: Technological developments in instrumentation such as the invention of the pneumatic trough for collecting gases over water, and later mercury, enabled researchers to discover a series of new gases in the period c. 1755–c. 1775, and to find out the properties of these gases. These discoveries concerning new gases were the trigger of the chemical revolution. This is an example of what I have called the 'tech first' pattern. In Sect. 4 will give an example of the 'tech last' pattern.

[7] This rather Kuhnian formulation has been called into question by two very interesting recent studies of the Chemical Revolution, namely Holmes [14] and Chang [3]. Holmes thinks that Priestley was not defending the old phlogiston theory, but rather a new phlogiston theory of his own invention, while Chang, who defends a pluralist view of science, argues that "phlogiston should have lived". These matters, however, relate more to the development and conclusion of the Chemical Revolution than to its beginning which is the subject of this paper.

4 Tech Last, and the Germ Theory of Disease

The tech last pattern is quite different from the tech first pattern. A tech last revolution is not preceded by developments in technology, which produce new instruments and a build-up of new discoveries. What precedes the revolution is rather a major unsolved practical problem. Attempts to solve this problem within the current paradigm have failed, and this encourages some researchers to seek a solution to the problem by changing the paradigm. Note, however, that the failure to obtain a solution to the practical problem may not be an anomaly in the current paradigm, because there may be no guarantee that the problem is in fact soluble.

I will now illustrate these features of the tech last pattern by an example drawn from the history of medicine. One of the biggest revolutions in medicine started about 1865 and had largely succeeded by about 1885. This revolution established the germ theory of disease as a new paradigm for medicine, and brought antisepsis into the practice of surgery.

Nowadays we are all completely familiar with the idea that a wide range of diseases are caused by bacteria or viruses. It is therefore rather surprising to learn that such a germ theory of disease was, apart from a few precursors, only introduced around 1865, and only came to be generally accepted by the medical profession around 1885. The germ theory of disease was first successfully used to explain two diseases—wound suppuration or sepsis, and anthrax. Anthrax is a disease of both humans and cattle. A French doctor Casimir Davaine (1812–1882) suggested in 1863 that it was caused by microbes, which he called *bacteridia*. His view was initially criticized and rejected, but came to be accepted much later (about 1881) owing to further work on anthrax by Koch and Pasteur. The other pioneer of the germ theory of disease was Joseph Lister. He used the germ theory to explain wound suppuration, and was more successful than Davaine, as he managed to get his view accepted by the medical community. I will now analyse what led Lister to begin his revolution in medicine.

Joseph Lister (1827–1912) was elected to a Fellowship of the Royal Society in 1860, and appointed as Regius Professor of Surgery at Glasgow the same year. It was while at Glasgow that, in 1865, he introduced his antiseptic system of surgery.[8]

To understand Lister's innovation, it is necessary to know something about surgery as it was practised in the 1860s and 1870s. Anaesthesia for surgery had been introduced in 1846. Indeed Lister, as a medical student at University College London, had attended the first operation to be carried out in Britain using anaesthetics. So the horrors of pre-anaesthetic surgery were over, but surgery was still in a very unsatisfactory state. The main problem was that, after any operation, the wounds instead of healing might become severely inflamed and then turn septic producing pus. Such sepsis, or suppuration, often had fatal results, but it was not

---

[8] In my account of Lister and his work, I have found Godlee [10] and Harding Rains [12] very useful.

known why it occurred. Harding Rains gives the following vivid description of what could typically occur in a hospital in the 1860s [12, p. 46]:

> And why was there such a danger? We, today, know the answer – infection by bacteria – but Lister and his world did not. They saw the effects, not knowing the cause. They had to stand and see the skin and flesh around wounds become intensely red and hot and swollen. This suppuration, as they called it, would get worse, the skin turning black with gangrene. Foul-smelling fluid and pus would run out of the patient due to the rottenness or putrefaction which seemed to be eating its way into the body. The body would become full of poison (septicaemia), causing shivering, high fever, wasting of the whole body, which would end in death. The whole state of affairs they called "sepsis". It was indeed dreadful sepsis.

What was the effect of all this on patient mortality? Mr. Erichsen, Professor of Clinical Surgery at University College London, published a booklet on the question: 'Hospitalism and the Causes of Death after Operations' in 1874. As regards death after amputations, he regards a mortality rate of 24–26 % as 'a very satisfactory result'. This 'very satisfactory result' occurred at University College Hospital. Similar levels were achieved at the Pennsylvania Hospital and the Massachusetts General Hospital in Boston. However the rate at the Edinburgh Infirmary was 43 %, at the Glasgow Infirmary 39 %, in Paris around 60 %, in Zurich 46 %. In military field hospitals the mortality rate for amputations was 75–90 %. A few years earlier in 1871, Sir James Simpson, who had introduced chloroform as an anaesthetic, published a series of articles on 'Hospitalism', in which he made the famous claim that "the man laid on the operating-table in one of our surgical hospitals is exposed to more chances of death than the English soldier on the field of Waterloo."

Harding Rains is quite correct to say that the cause of sepsis, or suppuration, was not known in the early 1860s. Nonetheless there were theories of sepsis at that time. Surgeons had to concern themselves a good deal with broken bones. Such fractures were divided into *simple* and *compound*. In a simple fracture, the skin remained intact; whereas in a compound fracture the bone penetrated the skin. The prognosis in the two cases was very different. Simple fractures could normally be set, and then healed up without any problems. Compound fractures, however, usually became septic so that amputation of the limb could not be avoided. These facts suggested that exposure to air was one of the causes of sepsis, and one theory was that sepsis was brought about by oxidation. We know Lister's own views in the early 1860s because he wrote some notes on suppuration on 12 December 1861. These have survived, and in them Lister concludes that bodily fluids after they have been acted on by air acquire chemical properties which cause suppuration. This kind of approach was supported by the general background in science and medicine of that time. In medicine, it was believed that many diseases were caused by miasmas, or bad airs; while chemistry was one of the leading sciences of the time. All this suggested that suppuration was a chemical process initiated by exposure to air.

The only problem with this theory was that it seemed to make it almost impossible to prevent suppuration, since in most operations air could not be excluded from the wounds. Lister's two teachers—Liston and Syme—tried to devise ways of dressing wounds which would keep out air. Liston used water dressings, and Syme dry dressings. Lister tried both, but without much success. The problem of sepsis appeared to be insoluble. It was this situation that stimulated Lister to challenge the dominant miasmatic paradigm of disease. Lister had a colleague at Glasgow, Thomas Anderson, who was Professor of Chemistry. In 1865, Anderson suggested to Lister that he should read some interesting recent papers by a French chemist, Louis Pasteur, who had been doing research into fermentation. It was the study of Pasteur's work, which gave Lister the clue to solving the problem of sepsis.

Up to 1856, fermentation had been thought to be a chemical process – a theory, which went back to Lavoisier. However, in 1856, Pasteur who was then Professor of Chemistry at Lille did some research into the production of ethyl alcohol by fermenting beet sugar. Pasteur concluded that fermentation is not a chemical process, but is brought about by a micro-organism (yeast). He published his results in 1857 in his *Mémoire sur la fermentation appellée lactique*, and this led in the next few years to the majority of the scientific community coming to accept the microbiological theory of fermentation.

Another important result, which Pasteur had shown before 1865, was that the air is full of microbes, which float on dust particles.

We can now reconstruct how the study of Pasteur's writings influenced Lister. Lister's work on wound sepsis had suggested that air was an important factor in causing suppuration; but he still thought that this must be due to chemical changes produced by the air. On reading Pasteur, however, he rapidly reached the conclusion that it was not the air itself but micro-organisms contained in the air which were responsible for wound sepsis. This is how Lister himself puts it in his 1867 paper: On the Antiseptic Principle of the Practice of Surgery [22, p. 133]:

> To prevent the occurrence of suppuration with all its attendant risks was an object manifestly desirable, but till lately apparently unattainable, since it seemed hopeless to attempt to exclude oxygen which was universally regarded as the agent by which putrefaction was effected. But when it had been shown by the researches of Pasteur that the septic properties of the atmosphere depended not on the oxygen, or any gaseous constituent, but on minute organisms suspended in it, which owed their energy to their vitality, it occurred to me that decomposition in the injured part might be avoided without excluding the air, by applying as a dressing some material capable of destroying the life of the floating particles. Upon this principle I have based a practice of which I will now attempt to give a short account.

Lister's aim was to find some way of keeping airborne microbes out of wounds. In principle he could have done this by heating, filtration or the use of chemical antiseptics; but in practice he regarded the last option as the only feasible one. For his antiseptic, Lister chose carbolic acid, which had been used as a disinfectant in dealing with sewage in Carlisle, and was easily available. His first attempts at antiseptic surgery in March 1865 consisted of an operation on the wrist and a compound fracture. Both failed and this led Lister to further reflection and refinement of his technique. This resulted in his first striking success on 12 August 1865.

The patient was an eleven-year-old boy, James Greenlees, who was brought to Glasgow Infirmary with a compound fracture of the leg, caused when he was run over by a cart. The boy was anaesthetized with chloroform, and the wound washed out thoroughly with a solution of carbolic acid in linseed oil. It was then dressed with a mixture of putty and carbolic acid, the putty being used to hold the antiseptic in place. This dressing was extended some distance from the wound, and covered with tin foil to help prevent evaporation of the carbolic acid. Finally, the leg was splinted, with bandages to hold both splint and dressing in place.

Four days late the dressing was removed. The skin was very sore, but there was no sign of putrefaction. Normally suppuration would have begun by this time. So it was an encouraging sign. Lister repeated the dressing and waited 5 days. During this time, the boy's temperature remained normal, and he did not lose his appetite. When this dressing was removed, the skin around the wound had been burned by the carbolic acid. A final carbolic acid dressing was applied, and left for another 4 days. By this time, the wound had begun to heal, and Lister judged that the risk of suppuration had passed. He applied a water dressing, to give the skin burned by the acid a chance to heal as well. Six weeks and 2 days after his accident, James Greenlees left the hospital with two whole legs—a remarkable achievement for the time.

5 Combinations of Tech First and Tech Last

So far I have emphasised the difference between tech first and tech last scientific revolutions. The distinction between tech first and tech last is indeed an important one, but many scientific revolutions can be considered as involving both patterns. This is partly because scientific revolutions very often have different phases, and partly because it is often difficult to decide how exactly a scientific revolution should be characterised. I will now give an example of each of these two situations.

The Copernican revolution can be divided into at least two phases. The first begins with the publication in 1543 of Copernicus' *De Revolutionibus Orbium Caelestium*. I would classify this as a tech last beginning. By 1500, Europeans had discovered America and also a sea route to the East Indies. During Copernicus' lifetime regular long-distance seaborne trade was established between Europe and these regions. Now such trade provided a powerful stimulus to seek an improvement in navigation, and, since navigation was largely carried out by observing the heavens, this in turn produced a stimulus to create a better astronomy and more accurate astronomical tables. The first phase of the Copernican revolution was successful in this. In 1551, Copernicus' work was used by Reinhold to compile a new set of astronomical tables, known as the *Prutenic* tables, after Reinhold's patron the Duke of Prussia. Once Kepler had improved the Copernican theory by his new laws of planetary motion, he applied this new theory to compile a new set of astronomical tables. These were known as the *Rudolphine* tables after Kepler's patron, the Emperor Rudolph, and were published in 1627. As Kuhn says [16, p. 219]: "… the *Rudophine Tables* were clearly superior to all the astronomical tables in use before."

Despite these successes, the first phase of the revolution did not really bring the revolution to completion, because the work of both Copernicus and Kepler could be, and was, interpreted instrumentally as merely giving mathematical techniques for better computation of the position of heavenly bodies without implying anything about the real nature of the universe. To complete the revolution, a second phase was needed and this opened with the telescopic observations of Galileo.

As I have argued in Sect. 3, the second phase of the Copernican revolution was tech first—the new technology being the telescope. Thus the Copernican revolution as a whole involves both the tech last and the tech first pattern.

I now turn to my second example of a combination of tech first and tech last. In Sect. 4, I analysed a revolution in medicine as tech last. However, the developments of that period could perhaps be analysed not just as a revolution in medicine, but rather as a revolution in the bio-medical sciences of which the revolution in medicine was just a part.[9] On this analysis, we could take the revolution as beginning with Pasteur's new explanation of fermentation of 1857, this being the first instance of the replacement of a purely chemical explanation by one involving microbes. This general bio-medical revolution initiated by Pasteur is a tech first revolution, the new technology being the microscope. Microscopes had been much improved in the first half of the 19th century with the elimination of chromatic and spherical aberration. Interestingly Lister's father played an important part in these developments. By the 1850s good quality microscopes were readily available, and Pasteur, who had been trained as a chemist, started using one in his laboratory. This was an unusual step for as Debré says [6, p. 87]:

> … he broke new ground, or rather went against the customs and habits of the chemists by bringing in the microscope.
>
> To bring a microscope into a biochemical laboratory is a relatively incongruous thing to do, even today. At the time it was a quasi-revolutionary act, the more so since Pasteur did not know what he was looking for, and barely what he was looking at …

However, it was the discoveries which Pasteur made by looking through his microscope which led to his new theories of fermentation and putrefaction. So we can analyse the general bio-medical revolution which introduced microbe theories, as falling into two phases. The first phase, initiated by Pasteur was tech first, while the second begun by Lister was tech last.

6 Testing the Theory

It is always good to test a theory, and so, having proposed a theory as to why scientific revolutions begin, it is desirable to look for some way of testing this theory, a way which could be applied to any other theory of why scientific revolutions

[9] This possibility was suggested to me by some comments made by Avinash Puri on an earlier draft of this paper.

begin. Now questions have already been raised with regard to the Copernican revolution as to why it occurred when and where it did, that is to say in Europe in the 16th and 17th centuries. There are other times and places where this revolution might have occurred, but did not. In particular two candidates suggest themselves.

1. The first is the ancient Greek world in the period roughly 300 BC to 200 BC. We know that in this period (probably around 280 BC), Aristarchus proposed a theory of the cosmos in which the Earth rotated on its axis and moved in an annual orbit round the Sun. (I will refer to such a theory from now on as a heliocentric theory.) No exposition of this theory has survived, but we know of its existence from the writings of Archimedes and others (see [7, pp. 135–141] for details). Yet this theory did not lead to a scientific revolution in the ancient Greek world, whose last cosmological system, produced by Ptolemy in roughly the period 125 AD to 150 AD, is still geocentric.
2. The second is Asia (India or China) in the 16th and 17th centuries. At this point, Asia was just as wealthy and, in most respects, just as technologically advanced, if not more so, than Europe. Why then did the Copernican revolution occur in Europe rather than in India or China? This problem is sometimes known as the Needham problem, because it was first formulated by Needham in his 1956 paper, and then in Volume III [30] of his *Science and Civilization in China*. Needham concentrates on the comparison of Europe and China, and I will here follow him in this, though it would be interesting for further research to look at the cases of India, Japan, and other areas in Asia.

Needham's formulation of the problem is rather different from the one given here. In his 1956 paper, he does not refer to the Copernican revolution, and indeed the only mention of Copernicus in the paper is in connection with Copernicus' work on monetary reform [29, p. 341]. Instead, Needham speaks of [29, p. 329]: "The birth of the experimental-mathematical method, which appeared in almost perfect form in Galileo…".

The analysis of the Copernican revolution, given in Sect. 5, makes it fall into two phases. The first beginning in 1543 with the publication of Copernicus' *De Revolutionibus*, and the second in 1610 with Galileo's telescopic observations. This second phase was needed to carry the revolution through to a successful conclusion. So Galileo does figure in an important way in my account, but, here again, there is a difference from Needham, who does not mention Galileo's telescopic observations in his 1956 paper. Despite these differences with Needham himself, I will refer to the question of why the Copernican revolution occurred in Europe and not in China as the Needham problem.

I will now examine whether the theory proposed as to why scientific revolutions begin can provide a solution to the Aristarchus problem and the Needham problem. As was argued in earlier in the paper, the theory, as applied to the Copernican revolution, divides that revolution into two phases. The first phase, which began with the publication of Copernicus' *De Revolutionibus*, was tech last. A tech last revolution begins with a pressing practical problem, which is not easy to solve within the existing paradigm. This situation suggests to some researchers that it might be worth trying to change this paradigm, and that, within a new paradigm, the practical problem might be easier to solve. If the revolution is successful, the acceptance of the new paradigm enables new technologies to be developed which solve the practical problem. In the case of the Copernican revolution, the practical problem arose because of the establishment of long-distance, indeed global, seaborne trade. This produced a stimulus towards the development of better navigation, which in turn produced a stimulus to produce a better astronomy and more accurate astronomical tables.

The second phase of the Copernican revolution was tech first. Technological developments, which will be described below, led to the creation of the telescope, and, using the telescope, Galileo made a whole series of discoveries of new objects, whose existence had not previously been suspected. The existence and behaviour of these objects could be explained much more easily within the new heliocentric paradigm than within the old geocentric one.

Thus the key factors in starting the Copernican revolution and carrying it though to a successful conclusion were (a) the existence of long-distance (global) seaborne trade, and (b) the invention of the telescope. In the absence of these factors we would not expect a shift from the geocentric to the heliocentric view to occur.

Turning now to Aristarchus, we can easily see that neither of the factors (a) or (b) were present in the ancient Greek world in the period 300 BC to 200 BC, when Aristarchus produced his heliocentric theory. Sea borne trade certainly existed at that time, but it was confined to the Mediterranean Sea. Geocentric astronomy was quite adequate for navigating in that confined sea, and so there was no stimulus to produce a better, more accurate, astronomy. As for the telescope, not only did it not exist, but the basis for creating it, namely a developed glass manufacturing industry, did not exist either. Indeed it was not even known how to produce clear glass in the period 300 BC to 200 BC. To show how the technological situation changed between the time of Aristarchus and that of Galileo, I will now briefly sketch the development of the glass industry.

In the ancient world the glass industry was first developed under the Roman Empire, and, then after the interlude of the dark ages, it was further developed in the Feudal era. Neither the Roman nor the Feudal periods contain very exciting developments in theoretical science, but they do contain considerable advances in technology. To some extent, technology can develop independently of theoretical science, though there often comes a time when further technological advances do require, as a prerequisite, significant theoretical developments.

Turning now to the development of the glass industry, two important technological innovations occurred in the first century AD. The first was the introduction of the new technique of glass blowing. The second was the production of colourless or 'aqua' glass. These innovations enabled glass to be produced on a large scale, and glass became a common material in the Roman world. It was used for tableware, both drinking vessels and vessels to contain liquids. A great deal of Roman glass survives to this day.

With the fall of the Roman Empire in the West and the coming of the Dark Ages, much glass making disappeared. For example, glass had been produced on quite a considerable scale in Roman Britain, but the material almost disappeared during the Dark Ages in Britain. However, enough technical knowledge of how to make glass survived during the Dark Ages in places such as Torcello near Venice to make another advance possible once the Feudal system had become established throughout Europe.

With the revival of trade and industry in Western Europe in the 11th century, glass manufacture was developed and improved. It was possible by the 12th century to produce the stained-glass windows of the great cathedrals. In the 14th century, the use of glass windows in houses became common (cf. [2, p. 213]).

The Arabs made lenses in the 11th century. Someone in Western Europe (perhaps an Italian around 1286) had the idea of attaching two lenses to a frame to produce a pair of spectacles. Techniques had advanced to the point where cheap clear glass could be produced, and so the stage was set for the development from 1300 of a spectacle-making industry. This naturally required the trade of lens grinder.

One of the main European centres for both glass production and spectacle-making was Venice, where production was concentrated on the island of Murano. Already in 1301, there were guild regulations in Venice governing the sale of eyeglasses. Pictures of people reading with eyeglasses appear in Italy by 1352, and north of the Alps in Germany by 1403. Eyeglasses must have had a big impact on life, and, in particular, increased the productivity of many workers. Previous to their invention, anyone with defective vision would have been excluded from reading and writing, and also from carrying on many artisanal activities. The use of spectacles could open up these activities to more people and also allow many with initially sound vision to continue these activities for much longer, since vision usually begins to decline after about fifty.

Given these developments, it is really rather remarkable that the telescope did not appear until the 1600s. Since lenses for spectacles were being regularly produced, it needed only a little experimentation with combinations of two lenses to arrive at the telescope.

Galileo was in the Venetian republic at the time of his telescopic discoveries. Throughout his life, Galileo kept in touch with manufacturing industry and artisans. At the beginning of his *Two New Sciences* of 1638, he describes the work of artisans in the Venetian arsenal, and he must have been familiar with the Venetian glass industry as well. He was therefore in a good position, when he heard reports of the telescope from the Low Countries, to have one produced and appreciate its value. This account of the development of the glass industry shows how impossible it would have been for anyone around 200 BC to have invented a telescope.

Let us now turn to our second case, that of China and the Needham problem. As before we have to consider whether the two factors which we have analysed as being crucial to the beginning and eventual success of the Copernican revolution were present in China in the 16th and 17th centuries. The first of these factors was the existence of long-distance (global) seaborne trade. Now here developments in China are of considerable interest (see [9, pp. 398–405]). Gernet is of the opinion that at the beginning of the 15th century, the Chinese were technically superior to the Europeans as regards the capacity to make long voyages on the high seas [9, pp. 398–399]. Gernet is probably right about this, because two crucial improvements in European maritime technology, namely the compass and the sternpost rudder came from China (see [1, pp. 317–319]). At all events, seven big Chinese maritime expeditions took place in the years 1405–1433. These expeditions comprised several dozen very large junks, carrying over twenty thousand men, and were headed by the admiral Cheng Ho. They visited Java, Sumatra, India, Persia, Arabia and the East Coast of Africa. These expeditions might have been the preliminary to establishing large-scale seaborne trade between China and these regions; but this did not occur. The expeditions marked the highpoint of Chinese maritime involvement, and, after 1433, the emperor and his ruling circle decreed a policy of withdrawing from seaborne activities.

The contrast with Europe is remarkable. In 1434, the year after Cheng Ho's last expedition, the Portuguese rounded Cape Bojador in Western Africa. Their ships were unsuited to further expeditions to the South, but, after 1440, the Portuguese developed the lateen-rigged caravel, which enable them to continue their voyages further down the coast of Africa. In 1487, Bartholomew Diaz rounded the southern tip of Africa (the Cape of Good Hope), and then in 1497 Vasco da Gama sailed round Africa to India, establishing a sea route to the spice islands of the East Indies. Meantime in 1492, Columbus had discovered the West Indies. Moreover, these voyages of exploration and discovery were followed by the establishment of regular seaborne long-distance trade. According to Davis' analysis of Spanish trade in the 16th century, there were [5, p. 63]: "forty thousand tons of shipping going to America each year in the 1540s, rising to a peak of four times that level at the end of the century." In addition, of course, there was a considerable European eastern trade with Africa, India, and the East Indies. All this long-distance seaborne trade provided a powerful stimulus for the improvement of navigation, and hence of astronomy. This stimulus was missing in the land-based China of the same period.

Let us now turn to the second key factor in the Copernican revolution—the invention of the telescope. Here again, the results are interesting. Glass was produced in China, but its use was limited to making beads, and decorative plaques and disks. Moreover, archaeological evidence shows that such glass objects were rare. The Chinese, in contrast to the Romans, never used glass for tableware, and, as a result, there was no development of a major glass industry in China. The reason for this situation was that the Chinese concentrated on the production of ceramics and metal work, and in these areas they were much more advanced than the Europeans.

Without a basic glass industry, however, there was no development of lenses and eyeglasses in China. Eyeglasses are first mentioned in China in the 15th century, and it is stated that these were imported. There was thus no industrial basis for the invention and development of the telescope in China.

This then is my suggested solution to the Needham problem of why the Copernican revolution occurred in China and not in Europe. It is interesting to contrast this with the solution proposed by Needham himself. Needham writes [29, p. 343]:

> Interest in Nature was not enough, controlled experimentation was not enough, empirical induction was not enough, eclipse-prediction and calendar-calculation were not enough – all of these the Chinese had. Apparently a mercantile culture alone was able to do what agrarian bureaucratic civilization could not – bring to fusion point the formerly separated disciplines of mathematics and nature-knowledge.

This passage is by no means inconsistent with the view presented here. I have claimed that a key factor in the origin of the Copernican revolution was the establishment of long distance (global) seaborne trade. But why did the Europeans establish this kind of trade, but not the Chinese? As we have seen, the Chinese did have all the technical skills needed to make long-distance ocean voyages. So the technological factor is not crucial here, and we should look instead at social factors. Now the profits of long distance seaborne trade went mainly to the merchant class. So, perhaps, as Needham seems to suggest, the merchant class was unable to pursue its interests in that direction because of its subordination to the ruling scholar bureaucrats, whose interests were mainly landed. By contrast, the merchant classes may have been stronger in Europe relative to the landed classes. However, these conjectures take us away from questions about the development of science and into general questions concerning social and economic history. I will not therefore pursue them further here.

Acknowledgments Earlier drafts of this paper were read at the Integrated History and Philosophy of Science conference in Athens on 15 March 2012, and at the workshop on Heuristic Reasoning in La Sapienza, Rome on 14 June 2013. I am very grateful to those who made comments on those occasions, and who read and commented on earlier drafts of the paper. Particularly useful comments came from Alex Bellamy, Hasok Chang, Andrew Gregory, Avinash Puri, and Thomas Sturm. Some of these are detailed in the footnotes.

References

1. Bernal, J.D.: Science in History. The Emergence of Science, vol. 1. Pelican Edition 1969 (1954)
2. Bishop, M.: The Penguin Book of the Middle Ages (1971)
3. Chang, H.: Is Water H_2O? Evidence, Realism and Pluralism. Springer, Berlin (2012)
4. Copernicus, N.: De Revolutionibus Orbium Caelestium (trans: Dobson, J.F., Brodetsky, S., in Munitz, M.K. (ed.) Theories of the Universe, 1965). Free Press Paperback Edition, pp. 149–173 (1543)
5. Davis, R.: The Rise of the Atlantic Economies. Weidenfeld and Nicolson (1973)
6. Debré, P.: Louis Pasteur (trans: Forster, E., Hopkins, J., 1998) (1994)
7. Dreyer, J.L.E.: A History of Astronomy from Thales to Kepler. Dover Edition (revised with a foreword by Stahl, W.H.) (1953)
8. Galileo, G.: Sidereus Nuncius (trans: The Starry Messenger. In: Drake, S. (ed.) Discoveries and Opinions of Galileo, 1957). Anchor Books, pp. 21–58 (1610)
9. Gernet, J.: A History of Chinese Civilization (trans: Foster, J.R.) (1972)
10. Godlee S.R.: Lord Lister. Macmilllan (1917)
11. Grünbaum, A.: The Falsifiability of the Lorentz-Fitzgerald contraction hypothesis. Br. J. Philos. Sci. **10**, 48–50 (1959)
12. Harding Rains, A.J.: Joseph Lister and Antisepsis. Priory Press (1977)
13. Holton, G.: Einstein, Michelson, and the ‘Crucial’ experiment. Isis **60**(2), 133–197 (1969)

14. Holmes, F.L.: The "Revolution in Chemistry and Physics": overthrow of a reigning paradigm or competition between contemporary research programs? Isis **91**(4), 735–753 (2000)
15. Hoyningen-Huene, P.: Reconstructing Scientific Revolutions: Thomas S. Kuhn's Philosophy of Science. University of Chicago Press, Chicago (1993)
16. Kuhn, T.S.: The Copernican Revolution. Vintage Books, 1959 (1957)
17. Kuhn, T.S.: The Structure of Scientific Revolutions, 2nd ed. University of Chicago Press, Enlarged (1962/1970)
18. Lakatos, I.: Newton's Effect on Scientific Standards [Worrall, J., Currie, G. (eds.), 1978], vol. 1, pp. 193–222 (1963–4)
19. Lakatos, I.: Falsification and the Methodology of Scientific Research Programmes [Worrall, J., Currie, G. (eds.), 1978], vol. 1, pp. 8–101 (1970)
20. Lavoisier, A.-L.: Elements of Chemistry (trans: Kerr, R., Dover Edition, 1965) (1789)
21. Leicester, H.M.: The Historical Background of Chemistry. Dover Edition, 1971 (1956)
22. Lister, J.: On the antiseptic principle of the practice of surgery. (Reprinted in Pasteur, L., Lister, J.: Germ Theory and its Applications to Medicine and On the Antiseptic Principle of the Practice of Surgery.) Great Minds Series, Prometheus Books, 1996, pp. 133–144 (1867)
23. Lorentz, H.A.: Collected Papers, 9 Vols. Nijhoff, The Hague (1935–1939)
24. Lorentz, H.A.: The relative motion of the Earth and the ether. Collected Papers **4**, 219–223 (1892)
25. Lorentz, H.A.: Versuch einer Theorie der elektrischen und optischen Erscheinungen in bewegten Körpern. Collected Papers **5**, 1–137 (1895)
26. McKie, D.: Antoine Lavoisier. The father of modern chemistry. Victor Gollancz, London (1935)
27. McKie, D.: Antoine Lavoisier. Scientist, Economist, Social Reformer. Constable, London (1952)
28. Miller, D.C.: Ether-drift experiments at Mount Wilson. Science **61**(1590), 617–621 (1925)
29. Needham, J.: Mathematics and science in China and the West. Sci. Soc. **20**(4), 320–343 (1956)
30. Needham, J.: Science and Civilisation in China, vol. III. Cambridge University Press, Cambridge (1959)
31. Popper, K.R.: The Logic of Scientific Discovery. Sixth impression (revised) of the 1959 English translation, 1972. Hutchinson, London (1934)
32. Popper, K.R.: Testability and 'ad-Hocness' of the contraction hypothesis. Br. J. Philos. Sci. **10**, 50 (1959)
33. Worrall, J., Currie, G. (eds.): Imre Lakatos. Philosophical Papers, vol. 1, 2. Cambridge University Press (reprint of paperback edition, 1983) (1978)
34. Zahar, E.: Einstein's Revolution. A Study in Heuristic. Open Court (1989)

Withstanding Tensions: Scientific Disagreement and Epistemic Tolerance

Christian Straßer, Dunja Šešelja and Jan Willem Wieland

Abstract Many philosophers of science consider scientific disagreement to be a major promoter of scientific progress. However, we lack an account of the epistemically and heuristically appropriate response scientists should have towards opposing positions in peer disagreements. Even though some scientific pluralists have advocated a notion of tolerance, the implications of this notion for one's epistemic stance and, more generally, for the scientific practice have been insufficiently explicated in the literature. In this paper we explicate a characteristic tension in which disagreeing scientists are situated and on this basis we propose a notion of epistemic tolerance.

Keywords Disagreement · Tolerance · Pluralism · Rationality

1 Introduction

The topic of scientific disagreement is not new in the philosophical literature. Already Kuhn [27] emphasized its significance for the diversity of ongoing research, while the more recent literature in social epistemology has further elaborated on the epistemic value of dissent (e.g. [32, 46]). However, we lack an account of the epistemically appropriate response scientists should have towards opposing

C. Straßer (✉) · D. Šešelja
Centre for Logic and Philosophy of Science (CLPS), Ghent University, Ghent, Belgium
e-mail: Christian.Strasser@UGent.be

D. Šešelja
e-mail: Dunja.Seselja@UGent.be

J.W. Wieland
Faculty of Philosophy, VU University Amsterdam, Amsterdam, Netherlands
e-mail: jjwwieland@gmail.com

E. Ippoliti (ed.), *Heuristic Reasoning*, Studies in Applied Philosophy, Epistemology and Rational Ethics 16, DOI 10.1007/978-3-319-09159-4_6

positions in peer disagreements. Certain scholars arguing for a pluralist take on science (e.g. [5, 30]) introduced the notion of tolerance in order to answer this question. According to them, scientific inquiries are sometimes prematurely dismissed due to the lack of epistemic toleration. The stance of epistemic toleration is thus of special relevance for the heuristic appraisal of scientific ideas in the context in which scientists disagree.

Central to our analysis of this concept is a significant tension inherent in the notion of toleration. Suppose a person or group A has a commitment and feels justified in its stance on a given issue while a person or group B has a different, incompatible stance on the same issue. Suppose further that we hear of A and B to be *tolerant* towards each other. We can immediately notice that this implies the following tension. On the one hand, A does not give up on its commitment when tolerating B's stance and, on the other hand, both stances are incompatible. This opens an important question: What is the epistemic nature of this tension? Only having answered this question are we in a good position to develop an adequate notion of epistemic tolerance.

In order not to end up with an 'anything-goes'-pluralism, tolerance should not be granted willy-nilly and hence B's perspective and/or stance should exhibit certain properties that make it worthy of tolerance. We will argue that (at least in the context of scientific controversies) such a criterion concerns the question whether there are reasons to suppose that B's stance is the product of a rational deliberation. Hence, the focus point of our discussion on epistemic toleration are *rational disagreements* (RDs). After introducing RDs in Sect. 2, we will in Sect. 3 explain why it is very difficult to recognize that the opposing parties in a disagreement may be reasonable, both for the participants in a disagreement and for external observers such as historians or philosophers of science. We argue that nevertheless there are specific indices that suggest that a participant's stance may be the outcome of a rational deliberation process. This will lay the groundwork for answering the above question in Sect. 4 with the help of the distinction between object-level and higher-order evidence in the context of peer disagreements.

In view of this characterization we will develop and justify our notion of epistemic tolerance in Sects. 5 and 6. Not only is history of science filled with examples of a lack of epistemic toleration among scientists involved in controversies, but also monism is a frequent epistemic position among scientists (as [5, Chap. 5] argues). Many of these controversies are nowadays externally recognized as RDs, while internally they failed to be recognized as such. If a pluralists' plea for avoiding premature rejections of inquiries is to be taken seriously, an account of scientific tolerance that is essentially informed by the tensional situation of epistemic agents in RDs becomes an important normative addition to this plea.

We will argue that epistemic toleration is not exhausted by regarding the opponents as potentially rational but that it also comes with other duties on the side of the tolerator. It is our point that an appropriate response to the epistemic tension of RDs demands an active notion of tolerance according to which scientists critically interact with their opponents. We will show that thus conceived our account of epistemic tolerance preserves the epistemically and heuristically beneficial component of disagreement (namely, an argumentative challenge), while it prevents the negative one (namely, the danger of a dogmatic rejection of a promising inquiry). This active notion of tolerance has to be distinguished from tolerance-as-perspectivism or tolerance-as-indifference.

According to perspectivism, the reason why we should be tolerant toward differing views is that the other's view, just as our own, provides "just a perspective". However, with this move the essential positive commitment participants in a disagreement have towards their own stance is neglected: after all, they consider their stance as the correct/true/adequate one and not just as some perspective among many. As Bryant puts it: "the pluralist paradoxically calls for every perspective to abandon its perspective, thereby abolishing the very thing that makes a perspective a perspective."[1] This is where pluralism, as the view that various differing perspectives are equally admissible, is in danger of turning against itself by narrowing down the very entities it sets out to protect.[2] A similar argument applies to a pluralism based on the notion of indifference which demands from its disagreeing protagonists that they tolerate each other in the mere sense of being indifferent towards each other. Also here the positive commitment that is at the very core of the disagreement is not taken seriously since it implies a critical thrust towards the position of the opponent which is missing in the notion of tolerance based on indifference.

This criticism demands a more sophisticated notion of tolerance in view of which pluralism is less narrow but, as some may argue, more tension-loaded as well. Indeed, it is our stance that a philosophically informed and useful notion of pluralism needs to characterize and be characterized by the epistemic tension of RDs. In this paper we provide such a notion of tolerance.

In Sect. 7 we discuss some related topics. We answer the questions: where is our framework situated in view of the uniqueness thesis of rational belief formation, and how does it relate to the Millian notion of group inquiry and to Habermas' notion of communicative rationality? Moreover, we counter possible objections by Feldman and Fogelin against the interactive character of our notion of tolerance. The reader may proceed selectively in this section in view of her interests.

[1] This has been pointed out by Bryant [3]. Feldman [13] makes a similar observation which we will discuss in more detail in Sect. 7.4.

[2] Note that although this argument illustrates that pluralism as perspectivism attacks its own foundation, it is nevertheless different and logically independent from the more frequent argument: if pluralism is applied to itself as a position among many (incl. monism) it looses its normative force.

2 Rational Disagreement

Schematically, a rational (or reasonable) disagreement can be characterized as follows:

Rational Disagreement (RD)—schematic characterization

RD1 There is a disagreement concerning some issue A.[3]
RD2 There are reasons to suppose that the stance of each participant is the result of a rational deliberation.[4]

This characterization can be further specified in various ways.

Concerning RD1, one may demand that the participants are epistemic peers in various aspects: e.g., concerning their intellectual capacities and concerning their access to evidence. There are, in principle, no big problems with recognizing epistemic peerhood in scientific practice: as soon as scientists learn about each other's expertise (their research background, familiarity with the given topic, publication record, etc.), they can acknowledge each other as peers. Major scientific controversies are usually peer controversies, although an overly strict reading of "equal access to evidence" may render peerhood an idealization. After all, each scientist has a unique experiental history. This may result in a scientist weighing, processing, and reasoning on the basis of evidence different than other scientists. Although the intellectual capacities of the participants in a disagreement may be similar, scientists may still reason differently given the shared evidence.[5] Were we to impose an overly strict reading of epistemic peerhood our peers (were there any left) would hardly end up disagreeing about anything, especially not in a rational way.

Concerning RD2 we can distinguish between an *external* and an *internal* recognition of RDs. In the external reading an external observer, such as a historian studying the controversy, has reasons to suppose that the stances of the participants are the result of rational deliberation, while in the internal reading it is the participants themselves who have reasons to consider each other as potentially rational.

---

[3] There is an involved discussion in epistemology what a disagreement amounts to especially if it concerns expressions such as 'probably', deontic 'ought', 'might' etc. In this paper it shall suffice to stay on a more pre-analytic level since our main focus is on rationality in 'rational disagreement' and the epistemic tension of participants in a rational disagreement.

[4] Similarly, addressing the issue of a reasonable disagreement in politics McMahon [35] writes: Wherever we find political disagreement, the parties will typically be prepared to offer reasons for the positions they take. The different positions will, in this sense, be *reasoned.* But to assert that disagreement in a particular case is reasonable is to do more than acknowledge that the parties have reasons for the positions they take. It is to imply that at least two of the opposing positions could be supported by reasoning that is *fully competent.* (p. 1, italics added)

[5] E.g., Goldman [17] points out that "two agents can have different bodies of evidence that bear on norm correctness and are relevant to the reasonability of their respective attitudes." (p. 208) Differences in this kind of norm-evidence are a reason for him to suppose that RDs are possible even in situations of "material evidential equality".

Finally, we can distinguish different levels of severity of the disagreement:

1. *Strong disagreement*: (From my perspective) your stance is wrong. My criticism seems to render your stance inadmissible. We further distinguish:
   (a) there are nevertheless reasons/indices in view of which I can suppose that your stance may be the result of a rational deliberation
   (b) there are no(t enough) reasons/indices of RD.
2. *Weak disagreement*: I see where you're coming from, I prefer my stance but I consider your stance as admissible. For instance, I may argue that my research is the most promising and substantiated but your research is interesting enough and since there is always a level of uncertainty and openness our discipline is as a whole more robust if we follow both research paths.

Note that 1(a) and 2 are RDs in our sense.

Case 2 seems less problematic from the perspective of pluralism than case 1 since there are no obvious serious tensions between the opponents. However, in the history of the sciences we often find disagreements where the opponents have less charitable attitudes towards each other such as cases 1(a) and 1(b). As for 1(a), we will argue in the subsequent sections that sometimes it may be impossible to render the reasoning of the opponent as sufficiently rational (from my own perspective) to immediately consider her stance as admissible, but that, nevertheless, there are certain indices in view of which we can suppose that her stance may be the result of a rational deliberation. We have already argued in the introduction that a pluralism as mere perspectivism or indifference is inappropriate especially in face of such cases. There is an essential tension between, on the one hand, my positive commitment towards my own stance together with my criticism towards the other's stance that seems to render the latter irrational and inadmissible and, on the other hand, the fact that there are certain indices in view of which I can suppose that my opponent's stance may, after all, be rational.

Note that case 2 turns into case 1 as soon as the subject matter of the debate moves to the question which stance (mine or the opponent's) is preferable, since now it is impossible for me to say that the stance of the opponent (that his view is preferable) is admissible. This would essentially undermine my own preference. This can be rather consequential in case the two views e.g. compete for research grants or in a case the distribution of researchers from both camps in scientific gremia is in question.

As a final remark let us notice that RD is a concept that is difficult to digest, especially if the condition RD2 is interpreted strongly, such as

RD2′ There are sufficient reasons to believe that the stance of each participant is the result of a rational deliberation.

Take the case of an intrinsic RD with RD2′. An agent who believes in A, believes that she is rational in believing so, and at the same time believes that her opponent is rational in believing not-A based on the same evidence, seems to be in a conflicting state of mind. She would need to subscribe to an internalist version of extreme epistemic permissiveness to render the situation consistent: the view that "there are possible cases in which [she] rationally believes P and yet a belief in not-P based on [her] total evidence would be equally rational" [2].[6]

Our interpretation of RD2 is weaker than RD2′ since we do not require 'sufficient reasons'. Hence, we still need to spell out which kind of reasons we have in mind. This will be the topic of the next section. Subsequently, in Sect. 4 we will analyze the epistemic situation of an agent in a RD.

3 Concerning the Recognition of Rational Disagreement

McMahon [35] points out the following dilemmatic situation with regard to political disagreement:

> …One of the defining features of reasonable disagreement in politics is that the contending positions do not seem equally reasonable to the parties, despite the fact that all are reasoning competently. Opposing views seem mistaken. This means that the contending positions will not seem equally reasonable to the reader, or at least to a reader who is engaged with the issue. An engaged reader will be engaged on one of the competing sides, and regard the reasoning supporting opposing positions as mistaken. (p. 9)

This motivates a similar worry for scientific disagreements. A scientist, who is an expert in the given domain, will always take one of the sides in the debate, and be convinced that this side is the rational one (i.e. stems from a rational cognitive stance), while the other one is not or less so. The deeper the disagreement is, the more unjustified the opponents' side may seem to be. Thus, on the one hand, an expert will always already have taken a side in the debate. On the other hand, any 'outsider' who is not engaged in the debate will not have sufficient expertise to judge whether the disagreement is rational or not. As a result, one can never be sure that a certain ongoing disagreement is a rational one. Hence, we are confronted with the following dilemma: on the one hand, we have seen philosophers explicating why sometimes disagreements in science can be considered rational; on the other hand, it seems that exemplifying a RD in practice is impossible.

Is there a way out of this dilemma? The fact that the recognition of a RD seems to come with an inevitable degree of uncertainty does not mean there are no *indices* of RD, on the basis of which one can infer she *may* be engaged in one. In this section we will specify two types of such indices: *content-based* and *form-based* ones.

---

[6] The notion Extreme Epistemic Permissiveness is taken from Brueckner and Bundy [2]. Our specification is "internalist" in the sense that we consider the perspective of a participant in a disagreement (as opposed to the perspective of an external observer). Brueckner and Bundy contrast their notion to Epistemic Permissiveness (without "extreme") which is weaker since it also covers cases where one agent believes P and another one suspends judgment on P. For more on Permissiveness and its opponent, the Uniqueness Thesis, see Sect. 7.2.

3.1 Content-Based Indices

The first type of indices—*content-based indices*—is gained by analyzing the argumentative content of the dispute.

This type of analysis has usually been made from an external perspective, by historians and philosophers of science. For instance, Chang [5] has recently analyzed the disagreement between phlogistonists and oxygenists, which was at the core of the Chemical Revolution. By analyzing the rivaling systems of scientific practice, Chang locates different epistemic preferences between the two groups of scientists. First of all, he distinguishes problem-fields that each of the competing sides considered important. While some of the problems were acknowledged to be important by both sides, others were regarded important by only one of the sides (p. 20). Moreover, Chang argues that neither of the sides focused on more significant problems, since the whole set of problem-fields was scientifically worthwhile. In addition, both groups offered solutions to the shared set of problems. Nevertheless, they disagreed on the relative qualities of those solutions (p. 22). Next, Chang examines the epistemic values that scientists used when evaluating these qualities. He shows that while oxygenists placed a higher preference on simplicity, phlogistonists preferred completeness. They also applied one and the same value in different ways: "Both sides valued unity, and each side cited the kind of unity it was able to achieve as persuasive evidence in its own favor" (p. 28). Similarly, each side showed to value systematicity by accusing the other one for being arbitrary and haphazard (p. 29).

On the basis of such a reconstruction of the debate, Chang shows that the rivaling standpoints were incommensurable[7] in more than one respect, and that each side had a rationale, that is, good arguments underlying its stance. In this sense, we can say that their object-level disagreement could have been a rational one, or can be externally judged as rational:

> …there was a genuine methodological incommensurability between the two systems of chemistry. Joseph Priestley was not irrational or unreasonable in his resistance to Lavoisierian chemistry, nor was he alone. [5, p. xvii]

Chang's analysis is an external one. But what would happen if Chang's discussion was available to the participants in this debate? If Chang could have explained to Lavoisier and Priestley that their opposing stances are in fact incommensurable? Here we are facing another deep tension in the philosophy of science. We have, on the one hand, the *ecstatic view* according to which "we have all learned to see the duck and then the rabbit […]; it is harder to learn to flip the Necker cube back and forth, but most of us can do it […] we can even learn to think simultaneously in terms of different systems." (Chang, p. 265). Hence, according to the ecstatic view we can imagine that both sides would have agreed with Chang's points.

---

[7] The notion of incommensurability is here used in a Kuhnian sense [26, 28], for its detailed explication see [20].

Consequently, this would lead not only to a deepening of the understanding of the controversy (where each side recognizes their epistemic divergences), but possibly to its resolution. That is, provided that on the basis of the content-based indices both parties recognize the two opposing stances as equally justified but incommensurable, and thus neither as preferable. In this case, there may be no more reasons for them to disagree.

However, another option is that, although our scientist recognizes the reasoning of her opponent as being systematically and consistently based on a different set and/or on a different application of cognitive values and methodological standards which are deemed rational by her opponent, she nevertheless fails to identify these very values and standards as being reasonable.[8] This poses the question, whether in this case the content-based indices are really indices of her opponent's stance being the result of a rational deliberation. In some such cases the fact that there is a consistent and systematic set of values and methods that is acknowledged by a number of respected peers (see our form-based indices below) can undermine the trust she has in her own rationale and at the same time open the possibility of her opponents' stance to be rational. We will discuss this in more detail in Sect. 4 in the context of higher-order evidence.

On the other hand, we have the *enstatic view* according to which scientists are (sometimes) essentially locked in their worldview and making the switch from the duck to the rabbit would rather be a conversion process than a learned skill that can spontaneously be evoked. This approach is most prominently present in the work of Kuhn [26, 28].[9] Another example is Aristides Baltas' view, according to which a scientific controversy is a type of disagreement "that cannot be readily settled by resorting to the commonly accepted disciplinary canons for conducting the relevant inquiry, as these have been developed up to that time", since it occurs "when disagreeing scientists *do not share background 'assumptions'*" [1, p. 44].[10] Moreover:

> … in pursuing their different strategies, scientists are constrained by something they do not share and *are in no position to lay bare* on the table of discussion. Their disagreement amounts to a controversy because they debate an issue without rendering explicit the very factors whose *silent existence* precludes their all resorting at the same moment and in the same manner to the same set of criteria, norms, or canons.
>
> … the existence of a background "assumption", its particular "level", and the role it effectively plays at any juncture of a science's development *can be determined only ex post facto, after* the "assumption" has been disclosed *and* from the *new* vantage point created by the disclosure. (p. 45, italics added)

[8] See also Footnote 5, where it was pointed out that due to having different standards of norm correctness our agent may not accept the standards of her opponent as reasonable. The latter is due to the fact that in her own experience of doing research our agent may have been exposed to different "norm-evidence", i.e., evidence that supports the correctness of norms.

[9] See Šešelja and Straßer [43] for a discussion and criticism of this issue in Kuhn.

[10] Baltas' notion of background "assumptions" follows the Wittgensteinian idea of "quasi-logical, or rather *grammatical,* conditions allowing the concepts involved in the inquiry to make sense" (p. 41). He also distinguishes between different levels of background assumptions and on the basis of these between different types of scientific controversies.

Altogether this indicates that the rationale of the opponent or the content-based indices thereof may not be that transparent to participants in a RD. There are various reasons for this. We have seen that according to the enstatic view scientists may just be too involved in their cognitive perspective to understand the opponent. Similarly, with the proponents of bounded rationality we may argue that, even if in principle the ecstatic view is right, in real life situation with its limited resources such as time and experience, sometimes content-based indices may fall within my engaged and professionally formed epistemic blindspot.

Altogether, an internal recognition of a RD by means of the content-based indices may not always be possible. Hence, we are back at the dilemma brought up at the beginning of this section. But is there maybe another way for the participants in a debate to recognize that their respective opponents' stance may be the result of a rational deliberation? In the following section we offer an alternative.

3.2 Form-Based Indices

We have seen that content-based indices derive from specific rationales of the parties involved in a disagreement. Nevertheless, there is another type of indices scientists can use in addition, or in the lack of noticing the content-based ones, in order to recognize they may be involved in a RD.

McMahon makes a similar point with regard to the recognition of a reasonable disagreement in politics:

> Reflection on *the history of a dispute* can facilitate the identification of the underlying disagreement as reasonable. […] if a particular form of political disagreement […] generally *survives extended debate,* conducted in good faith, in the contexts where it arises, we will have some basis for confidence that the disagreement is reasonable. We will have *some basis for supposing* that the disagreement is grounded in competent reasoning carried out within the framework of different life experiences. ([35, p. 26], italics added)

Hence, the structural properties of a debate, that are manifested throughout its history, offer symptoms of RD. We will refer to these symptoms as *form-based indices* of RD since they do not refer to the argumentative content, but rather to the context or the format within which the debate is carried out. Note that McMahon clearly indicates that we are dealing with reasoning on the basis of uncertainty.

While McMahon speaks of political debates, similar structural properties can be found in cases of scientific disagreements that philosophers and historians of science have characterized as rational. Let us take a look at the most common ones.

First of all, *the length of a debate* plays a role of such a symptom. Scientific controversies usually last for an extended time period, that is, from a few decades to more than a century. The discussion that led to the Copernican revolution took a whole century, throughout which time both, the proponents of the geocentric view and those of the heliocentric view offered valuable arguments for retaining their respective positions [25, p. 26]. If we consider methodological controversies as a sub-class of the scientific ones, Laudan's example of the debate about the rule of "predesignation" illustrates an even longer one. This debate begun in the nineteenth century and is still ongoing [31, p. 35].[11]

Let us notice some other indicators of RD which Laudan offers in this example, and which are common not only for methodological but also for the object-level RDs. He writes: "A host of prominent thinkers have been arrayed on each side of this issue (Whewell, Pierce, and Popper for predesignation; Mill and Keynes, among others, against it)." (ibid., p. 36). This sentence points to additional two symptoms of RD: the *expertise* of the participant in the given domain, and the *number of experts* being arrayed on each side of the dispute. On the one hand, one's expertise is essential in allowing for the framework of epistemic trust. On the other hand, the more experts there are on each side of the dispute, the more convincing it becomes to think that each side has good epistemic reasons to argue for its stance.

Of course, each of these formal indices is only a symptom and as such insufficient to provide certainty of a RD. Together they function in a gradual manner, so that the more of them are fulfilled, and to a higher degree each is fulfilled, the more confidently can a participant in a debate conclude that the disagreement at hand may be rational. Nevertheless, it is of importance for the argument we will present in the remainder of this paper to notice that such a conclusion can only come with a degree of uncertainty.

4 The Tensional Character of Rational Disagreements

4.1 Object-Level and Higher-Level Considerations

We have started with a situation in which two groups of scientists disagree. For instance, despite the fact that they share a similar body of evidence they come to different conclusions concerning e.g., what hypothesis or scientific model explains the evidence in the best way. In Sect. 3 we started with observing the inherent difficulties participants in apparent RDs have in recognizing each other's stances as being reasonable. Therefore we proposed different types of indices that allow an agent to conclude that her opponent's stance may be the outcome of a rational deliberation.

Now we have to ask in what epistemic situation our agent is, after being equipped with considerations informed by these indices. In a sense these considerations bring something new on the table: the evidence base of our agent is enriched. However, the evidence provided by the indices is different from the evidence that informs our agent's conviction in A. In what follows it is useful to

[11] According to the rule of predesignation a hypothesis is tested only by the new predictions drawn from it and not by its ability to explain—ex post hoc—what was already known.

distinguish between object-level and higher-order evidence. Object-level evidence stems from the very domain our scientists are reasoning about: measurements, experiments, observations, “data”, etc. To this evidence reasoning forms such as induction, abduction, inference to the best explanation are applied in order to draw conclusions, to generate and develop hypothesis, etc. The outcome of this type of reasoning may be called object-level considerations. Most parts of scientific debates are usually carried out by means of object-level considerations, i.e., arguments that are based on object-level evidence. Often, for instance, a scientist may forward an argumentative undercut against the inference of another scientist from a certain experimental outcome. E.g., she may criticize her opponent’s conclusion for not being warranted since it is based on a certain assumption that cannot be upheld due to a specific observation. Her opponent may defend himself by calling upon yet other “data”.

Additionally, our scientists are sometimes confronted with *higher-order evidence* which does not stem from the object-level domain in which a scientist works qua scientist.[12] An example of such an evidence would be the mere fact that the scientific peer disagrees with her. Another example would be that she reads a psychological study that states that scientists in her discipline are highly likely to develop certain cognitive biases. Unlike criticism based on object-level considerations such as undercutting arguments, higher-order evidence is not directed against any particular inference step our scientist made, i.e., against a particular relation between evidence and belief. It is one thing to undercut the belief that a certain object is red by pointing out that it is behind a red glass or in red light, and another thing to be confronted with the mere fact that another usually reliable person insists that the object is not red. The former argument points my attention to (object-level) evidence that was either previously unknown or not considered as relevant and which renders my inference erroneous. The latter confronts me with the fact that a similarly competent reasoner proportions her beliefs in a different way although no object-level considerations put into question the way I obtained my stance. Christensen states: “unlike ordinary undercutting evidence, [higher-order evidence] may leave intact the connections between evidence and conclusion. It’s just that the agent in question is placed in a position where she can’t trust her own appreciation of those connections.” Christensen [6, p. 198] notes that the trust is undermined even if our scientist’s stance seems to her to be correctly obtained in accordance with her epistemic and methodic ideals. Since all her object-level considerations seem to still fully support her stance, Christensen points out that she needs to “*put aside*” or “*bracket*” (ibid., p. 206) her original object-level reasons were she to take serious the higher-order evidence.

We will discuss the tension between object-level and higher-order consideration more below. Before that, let us notice that our indices of RD provide higher-order

[12] Higher-order evidence has been discussed under different names e.g. by Feldman [12] ("second-order evidence") and Kelly [23] ("higher-order evidence"). Our presentation is mostly inspired by Christensen [6].

evidence. In a sense they amplify the higher-order evidence provided by the mere fact that a scientist's peer disagrees with her. According to the form-based indices there may be a number of scientific authorities who for a longer period of time keep on disagreeing. This emphasizes the force against our scientist's trust towards her own stance. Analogously, identifying with the content-based indices that her opponents' reasoning is consistently based on different values or methodological commitments may undermine the trust she has towards her own axiological and methodological basis.

Note that the epistemic access to the opponent's rationale by means of the content- and form-based indices is indirect and therefore insufficient to declare the opponent to be rational. My own rationale is different from my opponent's and (factually) preferred to hers. Still, the indices provide me with higher-order considerations suggesting that her stance may be rational. For suppose I were certain about the rationality of her stance: then the disagreement would vanish since I would consider our stances on par in terms of justification. Hence, we have to deal with an epistemic *blindspot*: exactly the fact that the rationale of the opponent falls within my blindspot seems to be part of what may be taken to support me in insisting on my stance.

In sum, the situation is schematically depicted in Fig. 1. Object-level evidence (experiments, data, etc.) supports an agent object-level considerations giving rise to her stance A, e.g., her conviction in the strength of a given theory (this is the thick arrow from "object-level considerations" to the "⊕" on the left side). There may be some evidence that causes explanatory anomalies to the agent's theory, thus we also have a light arrow to the minus on the left. Since we are in a disagreement,

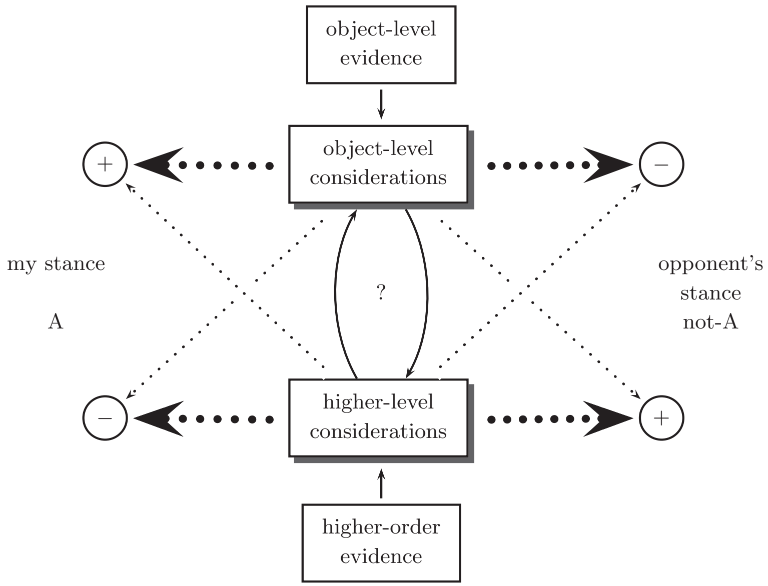


Fig. 1 The tension between object-level and higher-level considerations in peer disagreements

the object-level considerations support the agent's stance more than the opponent's stance, explaining the arrows to the "⊕" and "⊖" on the right. The higher-order considerations (e.g., the fact that peers disagree with her, the possibility of her being biased, etc.) stand in a tension with the former consideration since they seem to counter-weigh on them. For instance, the indices of rational disagreement provide some (indirect) support to the opponent's theory and undermine the agent's trust in her own theory. Similarly as in the case of the object-level considerations we do not exclude the possibility that there may be also higher-order considerations counter the opponent's theory (maybe there was a previous case of a hasty generalization in some experiment) or in support of her own theory: this explains the thin arrows from the higher order considerations to "⊖" on the right and "⊕" on the left. Nevertheless, typically there will be a tension between the support and dis-support in view of the object-level and higher-order considerations: while the object-level considerations support the agent's stance, the higher-order considerations counterbalance, and vice versa concerning the opponent's stance.

4.2 *Epistemic Responses to One's Own Stance in Peer Disagreement*

There are various possible responses to this situation in the literature on peer disagreement. They are mainly concerned with the question how the agent's *own stance* should be proportioned: the question whether or in what sense the higher-order considerations should influence her conviction in A that is formed on the basis of the positive support by means of her object-level considerations. In this paper we are also concerned with the dual question: the question of what is her stance *towards the opponent's stance* in view of this tension. Let us first review different stances concerning the first question. The most extreme view is that one type of considerations overshadows the other type. We can distinguish two variants:

1. Object-level considerations override the higher-level considerations: this means that an agent need not consider higher-order considerations when proportioning her stance, she need not 'bracket' the latter in Christensen's terms. She may remain fully confident in her stance. This option is taken by the so-called Steadfast Norm in the epistemology of peer disagreements. On a descriptive level we see this often with scientists: in view of peer disagreements they often tend to be unimpressed by the resistance of a group of peers. Indeed, the result is often even more than mere steadfastness: namely, an increased confidence in view of belief polarization.[13] On the normative level steadfastness can be critically associated with dogmatism: Were our agent to remain steadfast in her

[13] We will comment more on biases in Sect. 6.

stance this “can seem dogmatic […] if I concluded—on the basis of the very reasoning behind my own belief on the disputed matter—that my friend made the mistake in this case” [6, p. 206].[14] A less radical cousin of this approach is Conciliationism or the Extra Weight View “according to which one should give one’s own assessment more weight than the assessments of those one counts as peers.” [9, p. 176]. Although, according to this view, your confidence in your stance is lowered, you nevertheless remain steadfast since “you still think it more likely that you are right” (ibid., p. 176).

2. Higher-order considerations override the object-level considerations: this means that an agent needs to give up on her confidence in her stance in view of the higher-order considerations. For instance, such a view has been proposed in the literature on peer disagreement under the name Equal Weights View (see, e.g., [9]). In contrast to the steadfast norm that -as we have seen- seems more to conform to the default response of scientists, the requirement to give up on their confidence seems a rather difficult norm to follow for scientists in a peer disagreement. After all, “adjusting my belief to take account of my friend’s dissent can feel irrational […since it] requires the agent to put aside what still strike her as (…) clear reasons for her view” [6, p. 206].[15]

Kelly [23] takes a more moderate path: according to his Total Evidence View the question of how to proportion your doxastic stance in such a situation is different from case to case. While in some of the toy examples often referred to in the literature more extreme responses such as 1 and 2 above may be appropriate, in other cases we may tend more for a solution in between. However, the matter is contextual and not to be judged in principle from the philosophical armchair.

Similarly, Christensen concludes that, in general, “neither epistemic path [i.e., 1 or 2] […] seem[s] fully satisfactory” [6, p. 206]. This means that our agents are required to sustain the tension: both types of considerations are important and neither one should be buried under the weight of the other.

[14] Kelp and Douven [24] present a variation of the Steadfast Norm in which the permission to remain steadfast is temporarily limited in view of being associated with the epistemic duty to find reasons to "resist the peer's case in favour of his conclusion", or to find new supporting evidence, or to be able to "explain how one's peer could have become involved in error" (p. 105).

[15] The norms that are suggested by scholars in the epistemology of peer disagreement are often phrased in terms of "adjusting beliefs". This has been criticized by e.g., Elgin [10] as "wrong-headed" (p. 61) since "given a body of evidence, there is no choice about what to believe." (p. 60) Adopting the distinction between beliefs and acceptance from Cohen [7] she suggests to rephrase the debate in terms of acceptance rather than belief where "to accept that *p* is to adopt a policy of being willing to treat *p* as a premise in assertoric inference or as basis for action where our interests are cognitive." (p. 64) Related worries concerning the notion of belief and arguments in favor of various types of acceptance can be found in Elliott and Willmes [11]. We are sympathetic to this approach and our discussion is coherent with this rephrasing.

4.3 Shifting the Perspective Towards the Stance of the Opponent

As mentioned above, the epistemic paths we considered in the previous subsection are mainly concerned with the question how an agent should proportion her doxastic stance towards A in a peer disagreement on A. In this paper we shift the perspective and ask the question of how to deal with the opponent in such a situation. Clearly, in the Equal Weights View, I should consider my opponent's stance on par. The view seems to cohere well with the program of scientific pluralism: as long as the disagreeing scientists have sufficient reason to consider the respective theories as worthy of pursuit they will continue investigating their respective stances and moreover, they will have good reasons to critically correspond with each other. The reason is that their doxastic stances render neither theory preferable which motivates an incentive to keep up-to-date with developments in the other camp.[16] The situation is less clear with the other epistemic paths we considered above. Recall that the discussion above concerned the question what to do with my own stance in a peer disagreement: the Steadfastness approach [resp. (Extra Weights View)] claimed that there should be no (resp. slight) lowering of confidence in view of higher-order considerations, Kelly's total evidence view allowed for some influence of the latter (the degree depending on the concrete situation), while Christensen saw a dilemmatic situation in which both options—allowing for the higher-order considerations to influence my stance and not allowing them to do so—seem unsatisfactory. Neither option gives us a direct answer to the question what I should make of the stance of the opponent.

In what follows we will develop an answer to this question with the help of the notion "epistemic tolerance". We presuppose we have a peer disagreement among scientists. We do not presuppose the validity of any of the epistemic responses to the question of what to do with my own stance in a peer disagreement. Of course, adherents of the Equal Weights View will consider the situation of a peer disagreement as a context in which agents already violate their epistemic norm: after all, according to the Equal Weights View our agents should give up on their conviction on their respective stances and stop disagreeing. For adherents of the Equal Weights View our framework of epistemic toleration may thus be viewed as establishing a contrary-to-duty norm: a norm that is triggered in a situation in which agents violate the primary norm established by the Equal Weights View. For adherents of other epistemic stances our framework can be considered complementary in a more straight-forward way.

Let us sum up by listing different types of considerations that our agent in a RD faces.

PC Positive considerations in favor of her own stance informed by object-level evidence.

[16] The Equal Weights View has been criticized for the problems of rendering epistemic agents spineless and lacking self-trust. Elga [9] replies to this.

CC Critical considerations that our agent uses to criticize the alternative stances. It should be noted that **PC** and **CC** are usually not entirely distinct sets since my critical commitment towards alternative theories may very well be one of the reasons why I take my stance to be the best one and vice versa, merits of my stance can in a comparative manner be used to highlight (e.g., explanatory) shortcomings of alternative stances.

NC Negative considerations informed by higher-order evidence may put into question the trust our agent has in her own rationale and stance, and they additionally support the view that the stance of the other may be the result of a rational deliberation.

5 Epistemic Toleration

In the previous section we have explicated the three components (**PC**, **CC**, and **NC**) constituting the tension characterizing RDs. In view of this, let us now notice three specific aspects of a RD. On the one hand, in view of **PC** and **CC** a scientist is still involved in a disagreement: she is convinced of her own stance and finds the stance of her opponent epistemically *objectionable.* On the other hand, the indices pointing to the possibility that she may be involved in a RD imply that she should not dismiss the opponent's stance as merely irrational (**NC**). Hence, the opponent's stance is in a specific sense *admissible.*

These three components—objection, admission, and non-admission—structurally resemble the concept of toleration as it is discussed in the practical sphere (see [15]). Generally speaking, toleration is an attitude/stance/virtue that is adequate in specific contexts and that is thus associated with a conditional commitment: given that certain conditions hold you should tolerate or it is virtuous to tolerate. We will show that one of the specific properties of toleration as a conditional commitment is that its triggering conditions are in a tension and tolerance exactly concerns a (ethically, politically, epistemically, etc.) responsible way of dealing with this tension. Moreover, the action and/or attitude tolerance calls for is in an important sense a sustaining of the tension rather than a resolution or closure.

5.1 The Conditions of Toleration

Forst [15] characterizes three criteria that have to be considered when specifying toleration as a conditional norm:

Objection criterion. Tolerated beliefs, stances or practices are considered objectionable and in an important sense wrong or bad.

Acceptance criterion. There are positive reasons in favor of a tolerating attitude towards the belief/stance/practice of the other one that outweigh negative reasons in the relevant context.

Rejection criterion. The limits of toleration are to be specified in terms of criteria in view of which it is justified to reject the belief/stance/practice of the other one.

Clearly, the three types of criteria are inter-related and need to be weighed against each other in practical case. The norm of toleration is triggered in case both the objection criterion and the acceptance criterion are satisfied while the rejection criterion is not satisfied. Note that where the objection criterion fails, we consider the norm of toleration intuitively as vacuous since, where is nothing to object there is nothing to tolerate. There is obviously a tension between the acceptance and the objection criterion. That means that in the act or state of tolerating I still object to the tolerated stance. This will have important repercussions in the epistemic domain as we will see below.

In addition, the concept of toleration includes the following elements: the context of toleration, which regards the relation between the tolerator and the tolerated; the subjects of toleration (individuals, groups, the state); and the objects of toleration (beliefs, actions, practices).[17]

Let us then see how these three criteria, as well as the additional elements, of the practical notion of toleration can be translated into the context of science. In order to avoid ambiguities regarding "acceptance" and "rejection" which have a specific meaning in discussions on the methodology of science, we will use instead the terms "admission" and "disavowal".[18]

Objection criterion. The tolerated stance is considered objectionable and in an important sense epistemically problematic. Compare this to (**PC**) in Sect. 4.

Admission criterion. There are reasons—namely, the indices of RD—in view of which it would be wrong not to tolerate an objectionable stance. Compare this to (**NC**) in Sect. 4.

Disavowal criterion. There are reasons to consider the stance of the opponent as futile and thus they specify the limits of toleration. These reasons are strong enough to outweigh the admission criterion empirically backed up reasons to suppose bias or fraud on the side of the opposition, the refusal to take part in argumentative exchange, a systematic reluctance to put hypotheses under critical empirical tests, systematic self-immunization from empirical and argumentative scrutiny, etc.

Let us briefly specify the remaining elements of epistemic toleration: the context of toleration is in our case scientific practice; the subjects of toleration are individual scientists or groups of scientists; and the objects of toleration are doxastic

[17] Forst also includes the condition that toleration be practiced voluntarily.

[18] The objection and admission criterion make epistemic tolerance a second-order attitude (i.e., directed at first-order attitudes). For a similar emphasis on the second-order level, cf. Hazlett [19]. Hazlett suggests that in a peer disagreement about *H* you need not revise your first-order attitude towards *H*, but should suspend judgment on whether your and your peer's first-order attitudes towards H are reasonable (which he calls the attitude of 'intellectual humility').

attitudes or cognitive stances of scientists towards a certain scientific inquiry or some of its elements (hence, factual, methodological and axiological statements). In the next section we will propose the normative content of epistemic toleration. Before that, let us point out why introducing such an account is significant.

First of all, philosophers of science have been using the idea of toleration precisely to describe how scientists should regard the views they disagree with, but which seem not to be completely dismissible (e.g. [5, 29, 38]). Nevertheless, no detailed account of epistemic toleration has been offered yet. In contrast to the practical idea of toleration (as applied in ethics, politics, etc.), the toleration that stems from the above mentioned components is epistemic in character. First, each of the three components is based on epistemic considerations. Second, the normative consequences of this type of toleration are epistemic and methodological, rather than ethical. They concern one's doxastic attitude or an epistemic stance towards the views of the opponent, as well as the methodological steps that are to be based on this stance, rather than an ethical or a moral stance towards them.

But what does the notion of epistemic toleration add to the already existing normative concepts in epistemology and methodology of science? The primary significance of introducing the concept of epistemic toleration is that it is able to, on the one hand, appropriately characterize the above mentioned tension arising from the internal recognition of a RD, and on the other hand, to give epistemically fruitful guidelines to a scientist in this situation. Let us explain why. The specific property of the concept of toleration in general (that is, the one coming from social, ethical and political contexts) is the presence of the conflicting components of objection and admission. By translating this concept into the epistemic realm, we allow for a specific type of epistemic admissibility, namely the admissibility of views with which we disagree, but which we regard as possibly rational. This type of admissibility is not present in the standard toolkit of epistemology or scientific methodology, which consists of stances such as acceptance, rejection, pursuit, non-pursuit, consideration, entertainment, etc. This is due to the fact that none of these concepts captures the internal epistemic tension characteristic of a recognition of a RD as appropriately as the concept of epistemic toleration. Moreover, each of these stances may be itself a matter of RD (e.g. scientists may disagree on whether a given theory is worthy of acceptance, pursuit, etc.), where the stance of toleration is, in contrast to them, a second- order attitude (see Footnote 18).

5.2 The Normative Content of Epistemic Toleration

The key question now concerns the normative content of epistemic toleration: what does it means to deal in an epistemically responsible way with the tension between objection and admission? Given our previous discussion this means that tolerance concerns our readiness to act in an epistemically responsible way relative to both sides of the tensional state of RD: (**PC**)&(**CC**) and (**NC**). This way we obtain two norms which we will state now and justify in Sect. 6.

TOL1 In view of (**NC**) that is motivated by means of the indices showing that the stance of my opponent may be the result of a rational deliberation we have the duty to treat her stance in a charitable way as potentially rational, as potentially non-futile and thus potentially fertile.

TOL2 Moreover, we have

(a) the duty to consider the opponent's stance as a potentially serious competition and challenge to our own stance; and
(b) the duty to stay in critical correspondence with the opponent.[19]

Let us point out that the duties in **TOL2** come with certain practical implications. Most importantly, tolerating scientists are not to ignore the presence of the tolerated inquiries, or publicly dismiss them as epistemically futile. This holds for various aspects of scientific practice: from conference presentations, to publications, to teaching, to decisions made about research funds. For instance, they imply that one should mention the dissenting views when presenting the given domain, or that one should not obstruct the opponents' line of research (of course only so long as the rationality indications are in place), by disqualifying them in advance as inadequate applicants for research funds. Similarly, journal editors or conference organizers should not disqualify participants merely due to them belonging to an opposing camp of scientists.[20] Moreover, editors of scientific journals are, in view of our norms, motivated to invite and encourage discussions among members of rivaling camps, such as commentaries on each others papers in order to encourage toleration in scientific practice. Similarly, conference organizers are motivated to ensure that information flow between opposing camps is encouraged. For instance, they can try to prevent scheduling of talks that would keep each of the opposing camps in distinct sections. Instead, sections that include discussions among the rivaling groups, or key-note speeches followed by an opponent's commentary support the realization of the requirements of our norms.

[19] A comment is in place concerning the problem of intentional ignorance. Given a conditional norm an agent may try to avoid responsibility by means of trying to avoid the very knowledge of the fact that the triggering conditions of the conditional norm are met. In our case, an agent could try to ignore the recognition of the indices of RD and hence the recognition that the admission criterion is met. We take this problem to be a deep problem within meta-ethics which is not specific to our context and is in need of an independent unifying solution. Intuitively speaking, an agent who behaves in this way seems to be intolerant in some way. However, we decided to characterize our notion of epistemic tolerance in a non-circular way such that it does not also regulate the ways in which agents deal with its triggering conditions.

[20] Robin Warren and Barry Marshall document that this is precisely what happened to their submission of an abstract presenting their research on Helicobacter pylori, as the bacteria that are one of the major causes of peptic ulcer disease, to the Gastroenterology Society of Australia in 1983, when according to the dominant theory, bacteria were not among the possible etiological factors of this illness: "our abstract was not accepted, with the condolence letter from the secretary stating that, 'of 67 abstracts submitted we could only accept 56', thus our material must have been rated in the bottom 10 %!" [34, p. 184]. In 2005 Warren and Marshall were awarded the Nobel Prize in Physiology or Medicine for this discovery.

It is worth mentioning that our norms complement the account on epistemic responsibility such as the one proposed by Rolin [42] or the account of epistemic fairness proposed by Solomon [45]. On the one hand, the duty to remain in argumentative interaction with the opponent also derives from Rolin's idea of epistemic responsibility. Her norm requires from a scientist to be epistemically responsible towards those scientists that form her primary audience (such as a discipline-based community). That means that a scientist should provide sufficient evidence in support of those of her stances that have been challenged. On the other hand, the implications of our norms on those in a position to influence science policy (including the above mentioned editors of journals or conference organizers) correspond to Solomon's idea of epistemic fairness, which evokes the social side of epistemology [45, p. 148]. In view of Solomon's account of social empiricism, which aims at bringing about epistemic justice through science policy, it is important to discover biasing factors that may disadvantage a group of scientists. Our norms can help to this end.

6 Justifying Epistemic Toleration

We will now give a justification to the normative content of our notion of epistemic toleration.

6.1 Justifying TOL1

Why should we at all consider our opponent's stance as potentially non-futile although we clearly disagree with it ((**PC**)&(**CC**))? Why not favoring (**PC**)&(**CC**) over (**NC**) especially since the latter is anyway rather tentative: hence, why not dismissing my opponent's stance as irrational and thus futile?

To answer this question let us take another look at some apparent RDs in the history of the sciences. In his analysis of the Chemical Revolution Chang argues that the phlogistonist system was prematurely dismissed, which "allowed chemists to forget too easily about the need to explain things like the common properties of metals and their relation to electricity" [5, p. 285]. A similar case concerns Wegener's theory of the Continental Drift, which some earth scientists regarded as epistemically futile in spite of results that elaborated the originally provided evidence of drifting, and its underlying mechanism. As a result, even in the 1950s, long after Alexander du Toit and Arthur Holmes made major contributions to this theory, certain North American earth scientists remained to treat it as a worthless epistemic contender in their field (see [44]).

In both cases the epistemically irresponsible stance concerned the fact that, despite the availability of indices of RD, some scientists universalized their stance as the only possibly rational one on the matter which meant that they fully dismissed their opponents' stance. This lead to a premature dismissal of the opposing stance. In Sect. 4 we have argued that the presence of indices of RD provides higher-order evidence that epistemically counters this universalization. The presence of higher-order evidence has epistemic consequences: while it is not enough to (fully) undermine my positive commitment in my stance, it nevertheless gives support to the possibility that the opponent's stance is the result of a rational deliberation. Being ignorant towards this possibility may lead to premature and possibly dogmatic dismissals. Epistemic tolerance counter-weighs these tendencies.

An additional incentive to take the indices of RD seriously stems from the fact that cognitive biases may tarnish the judgment of scientists. They lead them to "see and remember a reality that is consistent with their beliefs and expectations, while motivational biases cause them to see what is consistent with their needs, wishes, and self-interest." [40, p. 649]. This leads to "important limitations in perspective taking" (ibid., pp. 641–642), to group polarization in which participants of a controversy have tendencies for "critically scrutinizing arguments and evidence that threaten [their] beliefs" (ibid., p. 637; see also [22]) and for 'rationalizing' their own stances, to overconfidence (see e.g., [21, p. 730ff)), to the fallacy of initiative [i.e., "a tendency to attribute less initiative and less imagination to the opponent than to oneself" (ibid., p. 730)], to unwarranted optimism, to loss aversion which results in a bias towards the status quo since "the advantages of any alternative to the status quo are weighted more heavily than its advantages, a strong bias in favor of the status quo is observed" (ibid., p. 738). Finally, biases go 'all the way up' since as Pronin et al. [40, p. 637] observe we have to deal with a "blindness about the role that … biases play in shaping our … views". Pronin et al. concluded that these biases may result in

> …a kind of worldview or lay epistemology that can appropriately be termed *naive realism.* That is, people persist in feeling that their own take on the world enjoys particular authenticity, and that other actors will, or at least should, share that take, if they are attentive, rational, and objective perceivers of reality and open-minded seekers of truth. [40, p. 646]

Indeed, as it has also been argued by philosophers of science, bias is a factor that may very well play a role in scientific practice (see e.g. [4, 8, 45, 48]).

It is important to notice that our charity principle **TOL1** characterizes the opponent's stance in terms of *epistemic fertility* (or *epistemic non-futility*) and potential rationality which is epistemically weaker than characterizing it as rational or ascribing rationality to it. That means that I could still suspect that my opponent is irrational, and at the same time uphold this charity principle. The fact that I consider her stance as potentially rational and fertile has repercussions in the way I engage with it argumentatively. It means that I treat seriously the possibility that its rationale falls into my epistemic blindspot and hence I should restrain from dismissing it as a futile inquiry. At the same time I am not obliged to naively ascribe rationality to it or to interpret it as rational as possible (by maybe violating its intended interpretation). This opens the possibility to criticize my opponent for instance in terms of rational shortcomings. In the next section we argue in view of **TOL2** that this is even my duty.

6.2 Justifying TOL2

Since with **TOL1** resp. (**NC**) I consider the opponent's stance as potentially rational, this immediately implies that I should treat the opponent's stance as a serious scientific candidate. Together with my positive commitment (**PC**) towards my own stance this implies that I should consider it as a potential challenge to my position. Hence, **TOL2.a**.

However, this has further repercussions for scientific discourse. As soon as I consider my opponent's stance as a serious challenge I have the responsibility to argumentatively interact with her, which is **TOL2.b**.

My positive commitment and the fact that I disagree with her commit me to both: to criticize her and to respond to criticism from her. This way I justify my own commitment and at the same time she will be challenged to become aware of possible shortcomings of her stance and in view of this to improve her stance.

The fact that I should consider such an argumentative challenge as potentially fruitful follows from **TOL1**: in view of treating her stance in a charitable way I take my opponent's stance as potentially rational and non-futile. This has, on the one hand, the consequence that I have reasons to expect that she will react to my criticism in a responsible way herself: she takes it seriously in her evaluation of her own stance and tries to tackle problems that I point out. On the other hand, since despite all the criticism I have for her position, I still consider it to be a potentially serious competitor/alternative to my own stance, I should also take the challenges of my opponent seriously in order to corroborate my own position.

Finally, it is epistemically responsible to keep on critically challenging my opponent in view of the leap of faith that I granted concerning the non-futility of her stance. The granting of this charity is epistemically responsible only if re-evaluated and critically challenged in the further run.[21] This way I will qualitatively advance the debate by triggering a defense from my opponent in which way she will be forced to further corroborate her rationale. This may in the longer run either (i) lead to a further justification of my leap of faith: given that a significant group of highly skilled scientists argumentatively resists my continued criticism may be taken as yet another index of the fact that indeed their rationale is the result of a rational deliberation (which nevertheless may fall within my "rational blindspot"). Or (ii) this may lead to a rational closure of the debate since either I may be able to convince my opponents that there is indeed something wrong with their stance or vice versa. Or (iii) she may fail to offer further arguments, loose support from other scientists, and eventually bring her stance to the state in which I no longer see the indices of RD.

Laudan pointed out that the possibility of criticism is important since it represents a significant and often fruitful type of challenge in scientific debates. He states that sometimes our rational toolbox (what we consider to be factual knowledge, our methods, our goals) may turn out to be incoherent, in a state of

[21] See also our discussion in Sect. 7.4.

"disequilibrium" [31, p. 55]. Hence, axiological and methodological change may be informed by what we have learned along the historical path of doing science. He reviews various fruitful ways of challenging the rationale of the opponent's stance, such as "critical tools which we can utilize for the rational assessment of a group of cognitive aims or goals" (ibid., p. 50). Critical argumentative exchange is able to pinpoint these shortcomings and lead to improvement.[22] Similarly, McMahon states that "the clash of opposing arguments promotes sound reasoning. It exposes bias and other sources of incompetence in reasoning." [35, p. 26]. As much as we have to take seriously the possibility that the rationale of the opponent falls within our epistemic blindspot (**NC**), as seriously we have to take the possibility that our criticism is justified (**PC**)&(**CC**). Epistemic tolerance means taking both possibilities seriously.

Conceived in this way, epistemic tolerance indeed concerns the sustaining of the epistemic tension of RD by means of critical argumentative exchange and thus the preventing of a premature closure of the debate. Note that giving into either (**PC**)&(**CC**) or (**NC**) would prematurely close the debate. If the tension is resolved in favor of (**PC**) then the stance of the other is finally dismissed which is premature and epistemically unwarranted in view of (**NC**) and hence the fact that there are reasons to suppose that the rationale of my opponent falls within my epistemic blindspot. If, on the other hand, the tension is resolved in favor of (**NC**) then I ascribe rationality to the stance of the opponent and hence consider his stance as equally justified, which is premature in the sense of being epistemically unwarranted in view of my (**PC**).

Let us close this section with some words on biases. As Carrier [4] argues, scientific pluralism allows for opposing parties to keep on challenging one another, which reduces the impact of confirmation bias. Notable, it only does so in case our scientists are also open for challenge and do not merely have an incentive to challenge. This is promoted by epistemic tolerance, namely by **TOL2**: due to (**b**) our scientists stay in critical discourse, due to (**a**) they take each other as serious critics which counterbalances confirmation bias. However, we have to be careful. Pluralism seems prone to the 'false polarization effect':

> Both sides in a conflict believe that although their own views reflect the complexity, ambiguity, and contradictions of objective reality, the views of the other side have been dictated and distorted by ideology, self-interest, and other biases. These attributions in turn lead the conflicting partisans to see the other side as extreme, unreasonable, and unreachable…. The result is an overestimation of the relevant construal gap between the modal views of the two sides and an underestimation of the amount of common ground that could serve as a basis for conciliation and constructive action. [40, p. 651]

This tendency is counterbalanced by **TOL1** and **TOL2(a)**: by means of the former I take my opponent to be a reasonable critic while the latter promotes an incentive to put her criticism into a constructive and fruitful perspective.

---

[22] The reader finds various historical examples in Laudan [31].

7 Epistemic Toleration in Context

Let us now provide some context to our notions of epistemic tolerance. In Sect. 7.1 we relate our results to the Millian notion of group inquiry. In Sects. 7.2 and 7.3 we discuss and relate our framework to some more general themes concerning rationality: on the one hand the so-called Uniqueness Thesis as opposed to permissive notions of rationality and on the other hand Habermas' and Friedman's communicative rationality. Finally, in Sects. 7.4 and 7.5 we discuss some possible objections by Feldman and Fogelin. The reader is encouraged to proceed selectively according to her research interest in this section.

7.1 Millian Group Inquiry

First, the duties we have presented are in accordance with the tradition of fallibilism and critical rationalism, which goes back to John Stuart Mill and his essay *On Liberty* ([36], orig. published 1859).[23] Moffett [37] summarizes the Millian theory of rational group inquiry by means of the following two principles:

1. *Millian Platitude*: X's theoretical beliefs are fully, perhaps even adequately, justified at a given time t only if they hold up against the strongest counterarguments practically "available" at that time;
2. *Collective Criticism Condition*: X can be adequately justified in thinking that X's current theoretical beliefs hold up against the strongest counterarguments practically available at t only if there is free and open critical discussion of those beliefs amongst X's epistemic peers with whom X genuinely disagrees.

These two Millian principles pose a requirement for one to engage in a critical discussion with one's opponents, which requires for their ideas to be taken seriously. Moreover, they pose a requirement to engage in a critical challenge of their ideas, which simultaneously means a challenge of one's own ideas.

7.2 The Uniqueness Thesis and Permissiveness

Our framework does not presuppose any commitment to neither the so-called Uniqueness Thesis (UT) nor to a permissive view on rationality. According to UT

[23] Mill writes: "Complete liberty of contradicting and disproving our opinion is the very condition which justifies us in assuming its truth for the purposes of action; and on no other terms can a being with human faculties have any rational assurance of being right" (p. 15).
Nor is it enough that he should hear the arguments of adversaries from his own teachers, presented as they state them, and accompanied by what they offer as refutations. That is not the way to do justice to the arguments or bring them into real contact with his own mind. He must be able to hear them from persons who actually believe them, who defend them in earnest and do their very utmost for them. He must know them in their most plausible and persuasive form … (p. 26).

for a given body of evidence there is a unique rational stance that is appropriate (see [13]). A permissive notion of rationality allows for various different rational stances in view of the same body of evidence.[24] If UT is true, RDs in the strict sense (compare RD2′ in Sect. 2) are impossible. Now, our notion of RD is weaker (see our RD2): it only requires that the object-level considerations are in a tension with higher-order considerations that are informed by the indices of RD introduced in Sect. 3. Such a situation does not require permissiveness and is thus fully consistent with UT. Of course, in view of UT and opposite to permissiveness, any consideration that my opponent may be rational is at the same time a counter-consideration to my conviction that I am rational. As a matter of fact, we cannot be both rational according to UT (i.e., as long as we share the same (relevant) body of evidence). But this does not pose an argument in favor of the impossibility of the type of disagreements we consider under the name RD. Our notion of RD and the subsequently developed notion of epistemic tolerance is perfectly consistent with the supposition that maximally one of the disagreeing agents is rational just as it is perfectly consistent with the supposition that maximally one of our agents is right about the matter of fact they are disagreeing about.

It is true, however, that under a permissive perspective a scientist's incentive may be stronger to be tolerating towards an opposing scientist. The reason is that under the permissive perspective she can grant that the other one is reasonable without at the same time undermining the reasonableness of her own stance. Thus, a tolerating stance of a scientist with a permissive take on rationality is less tensional than a tolerating stance of an adherent of UT.

In view of this it is worth pointing out that especially in the philosophy of science there is a traditional resistance towards UT. Let us close this section with reviewing some of the positions.

Laudan's 'reticulated model of rationality' is explicitly contrasted with the Leibnizian ideal according to which "all disputes about matters of fact can be resolved by invoking appropriate rules of evidence" [31, p. 5]. His 'reticulated model' of scientific rationality allows for a mutual adjustment and mutual justification to occur among all three levels of scientific commitment (ibid., p. 62): the factual, the methodological and the axiological (concerning the goals of science). For instance, the axiological level can be revised in view of new factual and methodological considerations. Moreover, the shared set of scientific goals does not uniquely determine the set of methodological rules, nor do shared methodological rules uniquely determine the factual (or theoretical) claims. Rationality concerns the question whether the three components are coherent with each other or form

[24] Permissiveness is systematically criticized by White [47]. There it is associated with Van Fraassen's epistemology, the Epistemic Conservatism of Harman and Lycan, and epistemologies that aim for reflective equilibria as Rawls or Goodman. A critical reply is to be found in Brueckner and Bundy [2].

an ‘equilibrium’ (see [31, p. 55]).[25] The rational equilibrium of an individual scientist may be informed by her unique experiental history and may hence differ among scientists[26]:

> It is frequently true, for instance, that scientists who are doing their best to follow appropriate norms of disinterestedness, objectivity, and rationality nonetheless find themselves led to very divergent conclusions. (ibid., p. 12)

Similarly, according to Rescher, the stances advocated by different researchers may be the product of altogether different rationales which are mutually incomparable in the sense that there is no absolute point of view from which one is “more rational” than the other. In other words, Rescher is a contextualist concerning rationality: “*what is rational* […] is by no means uniform from person to person but variable with situation and context.” ([41, p. 12], his emphasis). Rescher characterizes rational inquiry as a “most harmonious overall co-ordination between the information afforded us by our experience and our question-answering endeavours” (ibid., p. 65), as “effecting an appropriate alignment between our beliefs and the available evidence” (ibid, p. 8). The rationale of a researcher is not exhausted by its methods and values: what is crucial is the way they are applied and the way we put emphasis among them. What is at play are “people’s substantivity-laden epistemic standards (‘criteria of plausibility’)” (ibid., p. 37) which grow out of the individual experiental histories and give rise to different “cognitive perspectives” (ibid., p. 41). Rescher notes that constraints on evidence are not sufficient to ensure consensus, rather they are “useful devices for eliminating or reducing *mistakes*” while “their operation leaves ample scope for disagreement and diversity” (ibid., p. 38, italics in original).[27]

Recently, Kelly called into question the thesis of path-independence of rational belief acquisition according to which scientists who were exposed to the same evidence in a different order come to the same rational stance. Kelly’s argument is informed by empirical observations that agents treat arguments against what they believe “with a greater measure of suspicion and […] subject it to closer scrutiny” [22, p. 617] compared to the supporting arguments. Moreover, the effort they put into searching for alternative hypotheses is disproportionate to the confidence they have in their own doxastic stance. He concludes that given this asymmetry in assessment, prior beliefs may well inform the way we update our beliefs in view of new evidence and when engaging in debates. Clearly, this violates path-independence at least understood in a narrow sense which is summarized by Kelly

[25] Laudan states: "But beyond demanding that our cognitive goals must reflect our best beliefs about what is and what is not possible, that our methods must stand in an appropriate relation to our goals, and that our implicit and explicit values must be synchronized, there is little more that the theory of rationality can demand." (ibid, p. 64).

[26] McMahon [35] makes a similar point: "since personal histories of problem solving differ, competently reasoning experts can disagree" (p. 12).

[27] Similarly, Moffett [37] argued for the possibility of RDs in view of underdetermination and epistemic conservatism.

in the Commutativity of Evidence Principle.[28] It is worth pointing out that in a broader sense Kelly takes the principle to hold: namely if evidence is understood in a more general way than just as "consist [ing] of things that it would be natural to call 'data'" (ibid., p. 627/628). If we additionally include "everything of which one is aware that makes a difference to what one is justified in believing" (ibid., p. 628) we obtain what Kelly dubs 'evidence in the broad sense'.[29] Now, when updating our beliefs, evidence in the broad sense such as informed by the critical scrutiny with which counter-arguments are evaluated can make a significant difference as well as "the space of alternative hypotheses of which one is aware" (ibid., p. 620). According to Kelly, for evidence in the broad sense commutativity stands. But this holds only in a rather abstract sense, since he immediately adds that in concreto "historical facts about when one acquires a given piece of evidence might make a *causal* difference to which body of total evidence one ultimately ends up with" (ibid., p. 628). Indeed, if evidence in a broad sense is in part dependent on prior beliefs, then the question which broad evidence we gain and in what order we gain it is in an important sense restricted.

7.3 Communicative Rationality

As shown before, our notion of epistemic tolerance is situated deeply within a discursive context and concerns the rationality of the involved parties. As such it has clear links with the notion of *communicative rationality* that has been introduced by Habermas and further elaborated by Friedman in the context of philosophy of science. In this section we will analyze this relationship. Habermas characterizes communicative rationality as follows:

> [It] carries connotations that ultimately trace back to the central experience of the non-coercively uniting, consensus creating power of argumentative speech, in which different participants overcome their initially subjective points of view, and, thanks to the commonality of reasonably motivated convictions, assure themselves simultaneously of the unity of the objective world and the intersubjectivity of their context of life. [18, p. 10]

RDs cause a crisis for the consensus creating power of communicative rationality thus conceived. In Friedman [16] the notion gets an interesting, maybe unexpected twist: he claims that there is a "(communatively) rational route leading [from an earlier framework] to the later framework" (p. 101) that is incommensurable with it. He argues that communicative rationality does not "require agreement on everything, or even on very much" (pp. 93–94): what is needed is "agreement on how to

[28] The principle reads as follows: "to the extent that what is reasonable for one to believe depends on one's total evidence, historical facts about the order in which that evidence is acquired make no difference to what it is reasonable for one to believe." (ibid., p. 616).

[29] See also Footnote 5 for Goldman's notion of norm-evidence which seems to fall under or at least complement this approach.

engage in rational deliberation" (p. 93), on a "space of reasons". Now, if we conceive of this space of reasons in a narrow sense, i.e., as the mathematical principles and coordinating principles [i.e., principles that "mediate between abstract mathematical structures and concrete physical phenomena" (p. 78)] constitutive of the respective theoretical framework a scientist works in, then this will hardly help in overcoming an incommensurable gap since the incommensurability exactly concerns these very principles.[30] In view of this, Friedman widens the space of reasons accordingly. What is important at this point is that "conceptually problematic philosophical themes become productively intertwined with relatively uncontroversial and unproblematic scientific accomplishments" (p. 107). These hybrid philosophico-empirical clusters become parts of the space of reasons and thus provide argumentative moves for the participants in the debate. The required agreement in the space of reasons is not committal towards these clusters. Rather, what is required is "a relatively stable consensus on what are the important contributions to the debate and, accordingly, on what moves and arguments must be taken seriously" (p. 107). This way, participants can agree that a new scientific framework is at least a "live option" without having further commitments to them.

Let us relate our notion of epistemic tolerance to communicative rationality. Note that as long as scientists caught in a controversy take as their standards of rationality the very principles on the basis of which they form their opposing points of view, they will necessarily take the opposing point of view to be unreasonable (see **PC** and **CC**). The fact that according to Friedman's notion of communicative rationality they widen the space of reasons in such a way that it allows for argumentative moves that run contrary to their constitutive principles (narrowly conceived) means that they take the argumentative line of their opponent to be the process of a rational deliberation. Only this way they can genuinely consider the stance of an opponent as a "live alternative". In this sense Friedman's notion is complementary to our demands **TOL1** and **TOL2**: scientists take their opponents as rational as long as their arguments stem from a shared space of reasons and they take their argumentative moves as a serious challenge.

However, what is underdeveloped in Friedman's account are the epistemic tensions created by moving from the space of reasons narrowly conceived to a broader space of reasons that is populated with controversal philosophico-empirical clusters. Friedman states that scientists in this transition are confronted with a "deeply problematic …state of inter-paradigmatic conceptual limbo" (p. 115). What communicative rationality provides is an account of what is an empirical reason or justification. It thus creates a space of reasons which provides "the kind of consensus…necessary for *understanding*" (p. 93). But it leaves "still room for

[30] Friedman points out that "from the point of view of the old constitutive framework [the new framework] is not even (empirically) possible." (p. 99). Note that, according to Friedman, "[t]he standards of communicative rationality are given by […] an empirical space of possibilities or space of reasons" (p. 93). This clearly indicates that Friedman is forced to move from a narrow conception of his space of reasons (and thus of communicative rationality) to a wider notion in order to bridge the gap between incommensurable frameworks.

doubt" (p. 94) since scientist may reasonably disagree whether an argument is decisive. What Friedman's notion of communicative rationality does not account for are the epistemic tensions they have to withstand when they 'tolerate' argumentative moves and stances as reasonable although they go against their primary convictions. This is where it needs to be complemented by a notion of epistemic tolerance. RD exactly marks the spot where the objectivity-constituting aspect of Habermas' communicative rationality gets into a crisis since scientists fail to "assure themselves … of the unity of the objective world". It seems indeed a far cry from Habermas' purely consensus-oriented force of communicative rationality to our tension-sustaining notion of tolerance, even with the detour via Friedman.

7.4 Contra Feldman: Tolerance as an Active Stance

As we just argued, there is a close relation between epistemic tolerance and critical challenges. Tolerating a certain view *H* means not only that one has a certain positive (**PC**) and critical commitment (**NC**) to *H*, but also that one *actively engages* in the discussion about *H*. Specifically, our notion of tolerance implies two duties: **TOL2(a)** and **TOL2(b)**.

Yet, that tolerance would imply these duties is controversial. Here is an expression of this worry (please note that the notion of tolerance mentioned differs from our proposal, as we shall explain below):

> Neither intolerance nor [tolerance] is an acceptable response to disagreement. Advocates of both tend to fail to take seriously the arguments for views opposed to their own [13, p. 178]

The thought is this. Suppose you conclude that *H* is false, while there is an indication that you are in a RD: your opponent accepts *H* and her stance may be the product of a rational deliberation. Now you have at least the following options: (i) you could be intolerant, i.e. reject *H* despite the indication that your opponent may be rational, (ii) you could be tolerant, i.e. adopt the paradoxical stance of rejecting *H* yourself, and accepting that your opponent is still entitled to *H*, or (iii), which is Feldman's own preferred position, you could suspend judgment on *H*, and stay neutral regarding the issue until further evidence comes along, or until further steps are taken in the debate.

According to Feldman, furthermore, the problem with options (i) and (ii) is that they fail to take seriously the debate about *H*. Option (i) fails because in this case your opponent has no impact on your view at all, despite the fact that she may be rational. Option (ii) fails according to Feldman in the opposite direction: in this case you have no impact on your opponent's view, despite the fact that you have all the reasons to reject it. Hence, intolerance would be inappropriate because it does not admit challenges to one's own view, and tolerance would be inappropriate because it does not admit challenges to one's opponent's view.

In steps, Feldman's worry about the attitude of tolerance is as follows: (1) Epistemic tolerance towards *H* means rejecting *H* yourself, but accepting that *H* is epistemically justified to your opponent. (2) Holding that *H* is epistemically justified to your opponent (while rejecting *H* yourself) implies that you are not willing to pose critical challenges to *H*. (3) Therefore: Epistemic tolerance excludes critical challenges. The crucial premise is (2). Why would it hold? According to Feldman, you are not willing to pose critical challenges to *H*, because of your attitude that your opponent is entitled to *H*. Indeed, the challenges seem pointless exactly because they have no impact. Or again: if you reject *H* yourself but still hold that others may accept *H*, then it does not make sense to convince them that *H* is problematic and incorrect.

Nevertheless, the worry just identified does not apply to *our* notion of epistemic tolerance. The main point is that, according to our proposal, epistemic tolerance is essentially tensional and unstable. That is, it consists, as we have argued, of two commitments, i.e. a positive one (to your stance on the matter which includes the critical commitment concerning *H*) and a negative one (of accepting that *H* may be epistemically justified to others, because *H* appears to be the product of a rational deliberation). Now, one cannot simply adopt such a tensional and unstable stance, that is, in a justified and responsible way, without actively engaging in the discussion about *H*. For unless both the positive and negative commitment are justified, something has to go.

Here is how it works. On the one hand, you are entitled to hold that *H* is potentially epistemically justified only so long as your opponent is able to hold up against your challenges to *H*. Thus: the negative commitment of tolerance (**NC**) implies that one must pose critical challenges to *H*. This is duty **TOL2.b**. Here, posing challenges is no longer pointless: you formulate challenges to *H* in order to check whether you can tolerate *H* and its proponents in the first place. On the other hand, you are entitled to reject *H* yourself only so long as you are able to hold up against the defense of *H* by your opponent. Thus: the positive commitment of tolerance (**PC**) implies that you must take your opponent's defense of *H* seriously. This is duty **TOL2.a**.

Hence, on this account, Feldman's worry does not apply: tolerance towards H is compatible with, and indeed entails, taking seriously the debate about *H*, and with actively engaging in it.

7.5 *Contra Fogelin: The Non-futility of Argumentative Challenge*

Fogelin [14] has posed a potential objection to our notion of epistemic tolerance as an active stance for which critical argumentative exchange is essential. In a case of what he dubs "deep disagreement" it is not only the case that a resolution by means of rational argumentation is impossible, but "the stronger claim [holds] that the conditions for argument do not exist" or are "undercut" and hence "deep disagreements cannot be resolved through argument" (p. 5). According to this line of thought, any attempt of critically exchanging arguments "becomes pointless" in order to resolve a controversy "since it makes an appeal to something that does not exist: a shared background of beliefs and preferences" (p. 5). This can be summarized in terms of the following regulative:

FR If you are in a deep disagreement, you should stop arguing with your opponent since it is futile.

Now, suppose I am caught in a scientific controversy and I have some indices of RD. According to Fogelin this may give me reasons to suppose that I may be in a deep disagreement and, furthermore, this means that I have reasons to suppose that critically challenging my opponent is futile. Isn't our account of epistemic tolerance which calls for critical exchange of arguments unreasonable in view of this argument?

Let us relate **FR** to the epistemic tension which we analyzed in Sect. 4. In order to take (**PC**) and (**CC**) seriously we have to take our criticism on the opponent's stance at face value. This commits us to challenge the opponent. Fogelin may respond that this commitment is counterbalanced by the insight (on basis of the indices of RD) that I may be involved in a deep disagreement (**NC**) which at the same time implies that it may be futile to argue with my opponent. However, it is important that there is a significant degree of uncertainty involved. Indeed, there is a specific indeterminacy involved in RDs which is responsible for this uncertainty of recognizing RDs.

To see this, consider what it would mean to be certain that you are in a RD. It implies to be certain that the argument of the opponent is the product of a rational deliberation. However, since you are in a rational *disagreement*, you still argumentatively disagree with the opponent. But what can this possibly mean? It means you have critical arguments pointing to a shortcoming in the stance of the opponent. This seems not to be possible as soon as you say that her stance is perfectly rational. Now, you may still prefer your stance for some extra-argumentative 'reasons', but there is hardly a disagreement left as soon as the stance of the opponent is definitely recognized as being rational.

The indeterminacy of RD implies that we are never sure whether the antecedent of **FD** (that we are in a deep disagreement) is met. Indeed, the situation we actually face is quite different: we are convinced of our own stance (and hence disagree with the opponent) (**PC**) and yet we have certain indices of RD (**NC**). Fogelin argues that argumentation cannot serve the function of resolution in deep disagreements. However, whether my arguments (also) serve the function of resolution is an open question which can only be answered in the future by actually challenging the opponent. In other words (borrowed from formal logic), the question whether a disagreement is not deep is semi-decidable: in case it is not deep either my arguments or my opponent's arguments will—in the long run—lead to a resolution of the debate (see (ii) and (iii) in Sect. 6.2), in case it is deep, no resolution will happen despite our continued arguing.

Moreover, my counter-arguments serve other functions beside resolution as well. We have seen (see (i) in Sect. 6.2) that by critically challenging my opponent I test and corroborate the robustness of the leap of faith I gave to my opponent's stance in **TOL1**. Moreover, as pointed out before, the history of science is full of examples where critical challenges were fruitful in the further corroboration of theories. This relates to a point made by Lugg [33] in a critical discussion of Fogelin: argumentation is not a static endeavor where we skillfully make use of a fixed set of resources. “Instead of thinking of shared beliefs as ‘a common court of appeal’, we should think of it as a product of discussion, argument, and debate” (p. 49). Indeed, science is moving forward by investigating our environment. In view of challenges new insights may be gathered for instance by means of experiments. Here, argument serves the role of guiding exploration and inquiry. Whether this leads to a resolution of the given controversy is an open question.

8 Coda: An Instrumentalist Objection?

Let us conclude this paper with an important tension that we did not mention hitherto and that is of a different nature. Concerning the information flow in science there are two opposing positions:

1. The first stance takes a transparent and rich information flow and thus challenges, criticism, etc. between two camps in a controversy as a catalyst for the forming of rebust and epistemically proportionate stances. Mill’s doctrine of rational group inquiry is one of the forerunners of this view and critical rationalism is one of the most articulated proponents nowadays.
2. This view is opposed with the stance that takes restrictions on the information flow and the (restricted) isolation of scientific paradigms to be catalysts of effective science (e.g. [49, 50]).

Obviously, our notion of epistemic toleration coheres very much with 1 and thus, proponents of 2 may be dissatisfied with our account. We cannot and will not try to resolve this long lasting (rational?) disagreement in philosophy in this paper. Some tentative comments are in place nevertheless.

While our main interest concerns the rationality of an individual scientist who is caught in a scientific controversy, the main interest of proponents of 2 concerns the conditions under which science operates most efficiently in view of specific goals. Two thoughts are important here. First, there is no guarantee that there is no trade-off between the two interests: i.e., it may very well be that a science in which individual scientists’ rationality is optimized (e.g., their beliefs are optimally proportionated to the given evidence, they are non-dogmatic, take criticism seriously, their methods are optimally motivated in view of their research histories, etc.) may very well not be a most efficient science, and vice versa. Hence, there may be a tension between individual rationality and instrumental (group-oriented) rationality in science (see also [50] which supports this view). Second, for philosophers of science with a normative thrust there is a choice to be made between individual and instrumental rationality and if there is really a trade-off between the two of them there is no hope that there is a way to satisfy both of them optimally. This choice seems to be essentially informed by the question of what is our ideal of science. To caricature it a bit and supposing for the moment that stance 2 is the most efficient: do we rather want a science where the scientists are individually rational but the scientific machinery may sometimes move a bit slowlier than optimal, or do we want a scientific machinery that performs most efficiently but where the scientists may sometimes put on blinkers which make them suboptimal viz. slightly dogmatic epistemic agents? But, of course, even the question whether stance 2 is optimal with respect to the efficiency of science is debated and e.g. many critical rationalists would differ.

Acknowledgments The research of this paper was supported by the Special Research Fund (BOF) Ghent University and the Research Foundation—Flanders (FWO)—for Christian Straßer and Jan Willem Wieland as FWO postdoctoral fellows, and for Dunja Šešelja as a BOF postdoctoral fellow.

References

1. Baltas, A.: Classifying scientific controversies. In: Machamer, P., Baltas, A. (2000)
2. Brueckner, A., Bundy, A.: On “epistemic permissiveness”. Synthese **188**, 165–177 (2012)
3. Bryant, L.R.: The conundrums of pluralism. Weblog post at http://larvalsubjects.wordpress.com/2013/05/21/the-conundrums-of-pluralism/ (2013)
4. Carrier, M.: Values and objectivity in science: value-ladenness, pluralism and the epistemic attitude. Sci. Educ. (2012). doi:10.1007/s11191-012-9481-5
5. Chang, H.: Is water H_2O? Evidence, pluralism and realism. Springer, Berlin (2012)
6. Christensen, D.: Higher-order evidence1. Philos. Phenomenol. Res. **81**, 185–215 (2010)
7. Cohen, L.J.: An Essay on Belief and Acceptance. Clarendon Press, Oxford (1992)
8. Douglas, H.E.: Science, Policy, and the Value-Free Ideal. University of Pittsburgh Press (2009)
9. Elga, A.: Reflection and disagreement. Noûs **41**, 478–502 (2007)
10. Elgin, C.: Persistent disagreement. In: Feldman and Warfield (eds), pp. 53–68 (2010)
11. Elliott, K., Willmes, D.: Cognitive attitudes and values in science. In: 2013 (Proceedings) Issue of Philosophy of Science (Forthcoming)
12. Feldman, R.: Respecting the evidence. Philos. Perspect. **19**, 95–119 (2005)
13. Feldman, R.: Reasonable religious disagreement. In: Anthony, L. (ed.) Philosphers Without Gods: Meditations on Atheism and Secular Life, pp. 194–218. Oxford University Press, New York (2007)
14. Fogelin, R.: The logic of deep disagreements. Informal Logic **7** (1985)
15. Forst, R.: The limits of toleration. Constellations **11**, 312–325 (2004)
16. Friedman, M.: Dynamics of Reason. CSLI Publications Stanford (2001)
17. Goldman, A.: Epistemic relativism and reasonable disagreement. Disagreement, 187–215 (2010)
18. Habermas, J.: The theory of communicative action, vol. i. Beacon, Boston (1984)
19. Hazlett, A.: Higher-order epistemic attitudes and intellectual humility. Episteme **9**, 205–223 (2012)
20. Hoyningen-Huene, P.: Reconstructing Scientific Revolutions: Thomas S. Kuhn’s Philosophy of Science. The University of Chicago Press, Chicago, London (1993)
21. Kahneman, D., Tversky, A.: Conflict resolution: a cognitive perspective. In: Preference, Belief and Similarity: Selected Writings. Amos Tversky, pp. 729–746. Cambridge University Press, Oxford (2004)

22. Kelly, T.: Disagreement, dogmatism, and belief polarization. J. Philos. **105**, 611–633 (2008)
23. Kelly, T.: Peer disagreement and higher order evidence. Soc. Epistemology Essent. Readings, pp. 183–217 (2010)
24. Kelp, C., Douven, I.: Sustaining a rational disagreement. In: EPSA Philosophy of Science: Amsterdam 2009, pp. 101–110 (2012)
25. Kitcher, P.: Patterns of scientific controversies. In: Machamer, P., Baltas, A. (2000)
26. Kuhn, T.: Structure of Scientific Revolutions, 3rd edn. The University of Chicago Press, Chicago (1962 [1996])
27. Kuhn, T.: The Essential Tension: Selected Studies in Scientific Tradition and Change. University of Chicago press, Chicago (1977)
28. Kuhn, T.: The Road Since Structure. University of Chicago Press, Chicago (2000)
29. Lacey, H.: Values and Objectivity in SCIENCE: The Current Controversy About Transgenic Crops. Lexington books, Lanham (2005)
30. Lacey, H.: Pluralismo metodológico, incomensurabilidade eo status científico do conhecimento tradicional. Scientiae Studia **10**, 425–454 (2012)
31. Laudan, L.: Science and Values. University of California Press (1984)
32. Longino, H.: The Fate of Knowledge. Princeton University Press, Princeton (2002)
33. Lugg, A.: Deep disagreement and informal logic: no cause for alarm. Informal Logic **8** (1986)
34. Marshall, B.J.: The discovery that helicobacter pylori, a spiral bacterium, caused peptic ulcer disease. In: Helicobacter pioneers: firsthand accounts from the scientists who discovered helicobacters, 1892–1982, pp. 165–202. Black-well Science Asia (2002)
35. McMahon, C.: Reasonable Disagreement. Cambridge University Press (2005)
36. Mill, J.S.: On Liberty and Other Essays. Digireads. Com (2010)
37. Moffett, M.: Reasonable disagreement and rational group inquiry. Episteme J. Soc. Epistemol. **4**, 352–367 (2007)
38. Perbal, L.: $G \times E$ interaction and pluralism in the postgenomic era. Biol. Theor. 1–9 (2013)
39. Peter Machamer, M.P., Baltas, A. (eds.): Scientific Controversies: Philosophical and Historical Perspectives. Oxford University Press, New York, Oxford (2000)
40. Pronin, E., Puccio, C., Ross, L.: Understanding misunderstanding: social psychological perspectives. In: Heuristics and Biases: The Psychology of Intuitive Judgment, pp. 636–665. Cambridge University Press (2002)
41. Rescher, N.: Pluralism: against the demand for consensus. Clarendon Press Oxford (1993)
42. Rolin, K.: Diversity and dissent in the social sciences the case of organization studies. Philos. Soc. Sci. **41**, 470–494 (2011)
43. Šešelja, D., Straßer, C.: Kuhn and the question of pursuit worthiness. Topoi **32**, 9–19 (2012)
44. Šešelja, D., Weber, E.: Rationality and irrationality in the history of continental drift: was the hypothesis of continental drift worthy of pursuit? Stud. Hist. Philos. Sci. **43**, 147–159 (2012)
45. Solomon, M.: Social Empiricism. MIT press, Cambridge, Massachusetts (2001)
46. Solomon, M.: Groupthink versus the wisdom of crowds: the social epistemology of deliberation and dissent. South. J. Philos. **44**, 2842 (2006)
47. White, R.: Epistemic permissiveness. Philos. Perspect. **19**, 445–459 (2005)
48. Wilholt, T.: Bias and values in scientific research. Stud. Hist. Philos. Sci. **40**, 92–101 (2009)
49. Wray, K.B.: Kuhn's evolutionary social epistemology. Cambridge University Press (2011)
50. Zollman, K.J.S.: The epistemic benefit of transient diversity. Erkenntnis **72**, 17–35 (2010)

Heuristics as Methods: Validity, Reliability and Velocity

Anna Grandori

Abstract Research on innovative economic and organizational decision making processes is reviewed using epistemological criteria, showing that an array of effective, logically sound, and in that sense 'rational' heuristics can be specified—different from the repertory of 'behavioral', potentially 'biasing', heuristics usually considered. Two case studies of innovative decision making under uncertainty are then presented, on new product development (a major project for reducing traffic pollution) and entrepreneurial decision making (protocol analyses of financial angels' investing decisions); showing that the heuristics applied do resemble more the 'slow and safe' heuristics of scientific discovery, rather than the 'fast and frugal' heuristics of everyday life. A third case analyzes decision making on military flights, addressing the question of whether heuristics can be 'fast and rational' simultaneously. Results suggest that they can, and help in identifying the rather unexplored rational heuristics sustaining 'highly reliable' action under risk.

Keywords Heuristics · Epistemic rationality · Discovery · Reliability · Knowledge

1 Introduction

The paper is rooted in a long lasting research program in the logic of economic discovery and innovation conducted by the author [13, 14, 16], leveraging also on related research in fields such as entrepreneurial decision making [30], scientific discovery and technological innovation [3, 25] and strategic innovation [9].

Those works have explored a new field and contributed to the emergence of a new perspective, that deserves to be called a 'rational heuristic' approach: a model of decision that in the course of being 'heuristic' (based on 'methods for discovery') is also 'rational' (based on 'logically sound' methods for discovery). Those logically

A. Grandori (✉)
Bocconi University, Milan, Italy
e-mail: anna.grandori@unibocconi.it

E. Ippoliti (ed.), *Heuristic Reasoning*, Studies in Applied Philosophy, Epistemology and Rational Ethics 16, DOI 10.1007/978-3-319-09159-4_7

sound, 'rational heuristics' can therefore usefully complement the array of 'behavioral' (and potentially 'biasing') heuristics, provided by the behavioral perspective—the dominant approach to heuristic decision making in the economic and organizational field. Those heuristics have been reviewed and categorized in the format of a usable array or portfolio of effective decision methods [15]. The array of effective heuristics—thus intended—identified so far includes the following: *modeling* problems as cause-effect hypotheses on possible performance [1], rather than as gaps with respect to a given type/level of performance to be reached; crafting *robust* and *multipurposed* alternatives, and using multiple criteria for evaluation, rather than accept/eliminate alternatives against few and given parameters [14]; letting resources search for uses and means search for ends, rather than only the traditional reverse [29]; considering search stopping rules based on the reliability and validity of the problem model and the marginal contribution to knowledge of further research, and not only the marginal cost of search [4]; generating hypotheses by using theory and not only experience [10]; gathering information in a systematic and hypotheses driven way [11]; generating and testing hypotheses according to procedures that have the same logical structure of those applied in scientific discovery—e.g. through 'abduction' and 'falsification' [13, 14, 23, 28, 34].

The present paper presents recent field research conducted by the author and associates, aimed at showing the possibility, importance and performance implications of the use of those rational heuristics in different sub-fields of activity, and employing different research methods.

The different fields are all characterized by 'uncertainty' in an epistemic, Knightian sense: the possible actions and states of the world are in principle infinite, in number and kind; hence they cannot be completely listed, nor a number expressing a probability of occurrence can be rationally defined.

However, they differ in some interesting respects and distinctive types of 'complications', and therefore provide specific additions to our understanding of decision making under uncertainty: the development of new projects/products; the evaluation of new entrepreneurial projects; the undertaking of reliable action under risk.

The consequences of conditions of uncertainty in the sense of limited foresight and limited observability, have been traditionally theorized to be a 'reduction' of the quantity of information analyzed and a 'neglect' of possible contingencies [32], when not even a shift to random trial and error or to habit and routine-based action ([7, 8]). In other terms, as uncertainty grows, it is supposed that the rationality of decision making decreases.

However this is emphatically not, and luckily not, the approach taken by pilots, explorators, or entrepreneurs—nor should it be the approach of any wise decision maker under 'unforeseeable contingencies', especially if action is risky. The analyses of the 'rational heuristics' conducted hereon will also take some steps forward in our understanding of the proper decision procedures in situation involving not only uncertainty but also 'risk'—intended as the possibility of significant negative consequences—a type of decision still not fully understood in its rational logical structure [18, 29, 35, 36].

The term ‘rational heuristics’ may sound as an oxymoron after much use of the notion of heuristics in relation to biases [20] or at best to ‘simple’, ‘fast and frugal’ rules that are supposed to ‘make us smart’ [2, 12]. However, Simon’s original notion of ‘intendedly rational’ behavior, and even more the notion of heuristic in epistemology and philosophy of science, is that of a ‘method for discovering action’ [21, 22, 28]. As such, a heuristic can be a logically correct and rational method, as much as it can be biasing; it might reduce as much as it can increase effort (it can be more or less efficient). In discovery, it is obviously not possible to be ‘always right’, in the sense of not making any substantive error; but it is possible and important to avoid procedural errors, as measurement errors, biased inferences, accepting false hypotheses or rejecting true ones [33].

The reminder of this paper is dedicated to extend the empirical support and the portfolio of such rational heuristics in ‘economic’ decisions, broadly intended as decisions in which resources are scarce and performance important.

2 The ‘Slow and Savant Heuristics that Make Us Smart’

2.1 New Project Development

An interview based case study on the ‘Green Move Project’ (Municipality of Milan—Politecnico of Milan) allows to reconstruct the logic through which new solutions for reducing traffic and pollution were developed, leading to proposals such as car sharing in residential condominiums and in business firms.

The Green Move Project[1]

In 2011 the Regione Lombardia entrusted the Politecnico di Milano a project for improving mobility while saving energy and improving the environmental sustainability of traffic through car sharing. The current experiences in the town were limited to a traditional car sharing service and the maintenance of some stocks of cars for the public to use at stations or airports, both obtaining little response in that format.

In a first phase, a composite group of about 40 people—predominantly engineers with several specializations (economy, management, transport, math, environment, territory) and designers was formed.

The group started by analyzing the available literature and reports on car sharing experiences around the world. The result was to list approx 40 successful cases, and to list their features: 45 ‘attributes’ were thus identified, including for instance the area covered, the costs, the mode of booking and payment, the types of car, the services offered together with the car, etc.

One possible method for generating new solutions would have been to combine those attributes in all possible ways, in a blind trial and error process; this approach was discarded as it would have entailed to examine some 2^{45} combinations (in the simplest case of considering only two modes per each attribute).

---

[1] Source: Excerpts from an interview personally conducted by the author and published in Grandori [16].

> Then the group opted for a hypothesis based approach. The core hypothesis was that car sharing should have worked better if it solved problems of users through services attached to cars. Then a list of possible *functions* were identified in the considered experiences. The focus was here on detecting the main uses and users in a city environment. On the basis on the available data on the 40 experiences the purposes of moving, together with some anagraphic correlates (age, ownership of other cars, etc.) generated 8 functional profiles: e.g. 'from and to airports and stations'; 'shopping in the center'; 'moving on campus'; 'living the night'; … Cases were reassessed by combining their attributes with those functions, and it was discovered that not only a set of services were offered together with cars (corroborating the initial working hypothesis) but also that more than one service was offered in each experience, hence car sharing working solutions were multifunctional.
>
> In a second phase the project activities were leaded predominantly by designers. A workshop was organized for generating new solutions in a reasoned way. The 'reasoning' should achieve the generation of hypothetical alternatives with some chances of being workable in technical, environmental and social terms. The 'attributes' present in the literature/case reports were reorganized in 'dimensions': e.g. 'passive vs active' user; 'B2B' versus 'B2C'; energy served vs energy enabled; indifferentiated vs community based service. The 'creative' exercise was structured by combining those dimensions in pairs so as to generate quadrants, and by asking to generate a solution for each quadrant. In other terms, the possible features of an alternative were hypothesized first, and alternatives fitting those features constructed. Some 80 alternatives were generated in this way. For example, the combination between 'Community service' with 'Active user' and 'Passive user' generated, respectively, the idea of condominium car-sharing; and of personal company car, later developed in actually recommended solutions. A third promising idea emerged by 'B2C' with 'Energy served' approaches: the solution was called 'homo energeticus', and consisted in partnerships of energy firms offering the possibity to gain 'mobility credits', to be spent in car-sharing services, by adopting Energy saving behaviors.
>
> Notably this is a decision procedure incorporates principles of generating high quality solutions, by Pareto-improving them with respect to two criteria each time. Being partial solutions, developed with only some criteria in mind, those solutions could then be merged and combined to generate full solution to the initial problem: proposal for car sharing with high quality features on technical, environmental and social respects.
>
> The 'combinations' (called 'macro-alternatives' in the study) have been the 'Condosharing' (sharing a car in a condominium, for any purposes of the condomini); the 'Firmsharing' (offering to employees in a firm the use privately the firm cars out of working hours, when car would lay unutilized) and 'Networked city sharing' (the city infrastructures—such as hospitals, commercial centers, and municipality offices—are seen as traffic attracting poles, whereby they may be interested in offering car sharing services, eventually as a promotion/discount). It can be noticed that all these solutions leverage on the multifunctionality of resources (cars in this case) bringing about their fuller utilization, and on communities of users or offerers.

This process was conducted by professors and engineers, who are quite familiar with optimization principles and methods. Nevertheless, the problem is open, possible alternatives are infinite, and the problem is unique whereby there is no rational basis for assigning probabilities. Possible results can be predicted only thanks to causal hypotheses, not on the basis of frequencies or logical structure of the problem. Then they (rationally) shifted to a heuristic approach.

Some specific heuristics are also recognizable in the process, among those already identified in previous research. For example, the decision making team operationalized desired performance into multiple objectives/parameters (a '*multipurposedness*' heuristics). The advantages of multipurposedness in terms of flexibility, innovativeness, likelihood of finding solutions and risk reduction have been extensively pointed out [3, 5, 14]. The Green Move case also provides material for clarifying further heuristics for *integrating objectives*, and highlighting their importance.

Classic decision making models contemplate basically only two ways of integrating objectives: 'compensatory' and 'non compensatory' integration. The two approaches have been often considered as indicators of a maximizing/rational versus satisficing/behavioral decision making models respectively. The strategies typically classified under the compensatory umbrella include linear or additive difference models, whereas the ones seen as non-compensatory are lexicographic, conjunctive, disjunctive rules, as well as elimination by aspects (for a comprehensive review, see [27]).

However, in multidimensional complex problem solving, it is doubtful that compensatory approaches can be considered epistemically rational. In fact, much sensitive information may be lost in performing trade-offs between qualitatively different criteria for reducing them to a single utility numbered function [14, 31]. On the other side, if 'non compensatory' is taken to mean the use of multiple parameters as sequential or simultaneous acceptability thresholds, all interaction effects are lost and no improvement of solutions obtained.

The case clearly shows that a better heuristic is however possible. The procedure is going through various cycles of assessment, revising the weigh given to various parameters, taking into account the interactions among them, and generating hypotheses on which properties alternatives should have for scoring better on a wide as possible set of parameters. That's '*integration*', *not* '*compensation*'. It is the same logic that lead to Pareto-improvements in negotiations (where different objectives are hold by different actors). But such a 'logic of improvement' with respect to multiple parameters is obviously applicable also among different objectives held by a single actor, treating them as if they were incomparable utility functions.

Hence, a *logic of Pareto-improvement* (*and Nash-improvement*) is more epistemically rational than any 'non compensatory' rule on one side (explore the space above and beyond thresholds), and also provides an epistemically rational alternative to multi-criteria 'maximization' where the latter cannot be rationally applied (where no maximum can be defined, let alone the probability of attaining it; and utility cannot be compared inter-persons or inter-parameters) [14].

2.2 Entrepreneurial Decision Making

Studies on entrepreneurial cognition and decision making confirm the relevance of rational heuristics also in condition of even stronger uncertainty, where no problem is defined to start with.

In this domain, the heuristics identified as particularly important and effective include:

- given that the problem is not 'given', while resources are typically scarse, in economic innovative decision making, it is possible and effective to 'let resources search for uses' and functions, rather than searching for means/ resources to produce a specified end ([13, 14, [19] ,30];
- given that that the activity is new, relevant 'experience' may not be available, or would not be a valid knowledge base for the new activity; hence, generating hypotheses by using theory and not only experience [10];
- given that risk is involved, blind trial and error would not only be inefficient but highly destructive [1]; on the other side, 'complete scanning' of information would be endless (hence, also inefficient; [34]); the superior heuristics is, in business as in science, gathering information in a systematic and hypotheses driven way [11, 28

Recent research results on the decision processes of investors in new high tech entrepreneurial projects adds some quantitatively analyzed data in support of the proposition that the evaluation of innovative projects under uncertainty makes rich use of the above rational heuristics [17]. Decision protocols included sufficient information for measuring processes according to the intensity of use of the following heuristics: 'resources in search of uses'; 'scenario reasoning'; 'theoretical and causal modeling'; 'experience-based modeling and pattern-recognition'; use of 'multipurposedness' and 'objectives' integration' heuristics. In addition, the projects proposed to investors were selected so as to have independent information on their 'goodness': the success of these projects in generating funding from external investors (not part of the sample) during actual pitches was used to classify them as good versus bad opportunities and thus served as an indicator of investors' error rate. Therefore, initial empirical evidence relevant for our claim that these heuristics are effective and complementary, is provided by exploring the connections between the use of those heuristics and the chance of committing Type I or Type II errors (rejecting good projects or accepting bad ones).

Among the results we can signal that the highest success rates are associated to processes (a) using *both* 'resources in search of uses' *and* 'forecasts and scenario reasoning; (b) generating and evaluating hypotheses about the likelihood of venture success using *both* '*theory*' (of how human or physical materials behave) *and* '*experience*' (experimental knowledge of a field); (c) using *differentiated, multiple evaluation parameters*, but also *integrating* them (by modifying the weigh and kind of parameters for finding/developing alternatives).

The following protocol of decision process illustrates the use of some of these heuristics, namely:

- Multiple types and aspects of resources 'in search of uses' are considered.
- Causal models/reasons of why certain attributes of resources—teams rather than individual entrepreneurs; industry-specific competences and technologies—are employed;

- There is also use of Popperian heuristics such as asking disconfirming questions; looking for contradictions/check for consistency; generating reasons 'against' the theses under judgment.

How do angels decide?[2]

"Normally, you would look at the management team, and one of the key things is that it is not one individual, that's deadly, that's what happened to the one investment I did that went horribly wrong because a) one person can't do everything, and b) you do need different personalities. So the salesman has the vision, etc. And then you need the boring guy who gets it done, because oftentimes that charismatic CEO type of person actually does not focus on the detail, and the detail guy does not have the vision, so you would like to see at least 2 people. If there is technology involved, I love to see someone who understands it within the company, rather than subcontracting it, because I have seen that quite a lot.

..And obviously there is this classic thing you know, have they got experience of the relevant industry, I think you can get there in a number of ways. What really smart people do is if the management team doesn't have experience, they get non-executives or sort of mentoring type of people who are experienced in the field. So basically if you haven't got the experience, you do the effort and particularly young guys, actually recognize that they don't know some things; the company I have invested in recently, they have a three-man sort of wise monkeys, who know a lot about the business.

What are the other rules, ahm, look around the table, if they are all from banking, accountants, etc. get out! Because, where are the people who know the market? Somebody who knows this market should still be there, because if there is none of them, I mean I have been in banking and like they know nothing about the real world.. and if you are a lawyer or an accountant, you are almost as bad as me. So that's with reference to the management team…

Then the numbers always make me laugh. Some of them are just hysterical. So they expect in 3 years time, to be valued at 1.5 mln? So for the hell of it, it just sounds wrong, these numbers, there is no room for error in these forecasts. .. .Ahm, one of the things I want to know is how has the company been funded. So how much money has been put into it, have the founders put in any money, and ..And always the question that you ask is, if you are doing so well, why aren't you raising the money yourself? Once you get to 2 mln pre-money, you should have a reasonable investor rate, 250 k is sort of nothing, it is just a chunk really..they have predicted a 150 k profit for next year so some of it could be funded out of their forecast. If the amount of money being raised is relative to the valuation.

That's always a bit bothering me. Yeah, if the amount of money being raised relative to their valuation ends up ten times the amount …then it is a bit odd. Because if it is going that well, the existing guys should want to do the investment themselves."

[2] Source: Protocols of decision processes from an experiment conducted with real business angels, analyzing real business plans in a simulated laboratory setting. Original material gathered by Magdalena Cholakova in her Phd Thesis work, under the supervision of the author. For more details see Cholakova and Grandori [6].

3 The 'Fast and Rational Heuristics that Make Us Safe'

One limitation to the use of the above described rational heuristics could be that, precisely as they resemble the 'slow and savant' procedures applied in scientific discovery, they might not be applicable under time pressure. In addition, a certain rate of error may be acceptable in the areas considered so far, for the sake of learning; while there are contexts where errors should be avoided at all costs, as they have fatal and irreversible consequences. They include all areas of prompt intervention under high risk, such as natural hazards, medical first aid, flight and military operations. Therefore, the third decision area considered and data presented are on those, sometime called 'high reliability' action systems [17, 37]. There is considerable research on the organizational conditions that may enhance high reliability, but there is much less on the heuristics applied for generating reliable and valid action and approximating 'zero defect' or 'zero error' conditions.

A core question laying ahead therefore is: are there/which are the heuristics that in the course of being 'rational' (using valid and reliable knowledge, making action safe) are also 'fast'?

Some initial answers to this question are provided by an enquiry on the decision behavior of flight crews of the Aeronautica Militare.

The study is very recent and exploratory; the information used here is gathered through documents (internal publications), interviews to responsible and trainers of the Flight Safety unit, and observation of simulations of flight with 'variances' to be solved, used in pilot training.[3] The situation considered here is that of multicrew settings (two pilots and some flight assistants) which is the richest setting for extracting the relevant heuristic. The elements extracted from those sources are grouped here in three classes. The subsequent case history provides qualitative evidence on how these heuristics can work in practice.

3.1 *Prevention Heuristics*

A powerful alternative to prediction is prevention. However, a stark opposition among those two approaches is also unsatisfactory. The approach used in flights indicates that intelligent prevention involves a lot of prediction, albeit not of the kind that is usually assumed in decision theory.

There is no prediction of which single states of the word or contingencies will materialize and with which probability. Rather, possible 'contingencies' or

[3] Evidence has been gathered in the framework of a research collaboration agreement between CROMA (Bocconi Research Center on Organization and Management) and ISSV (Superior Institute for Flight Safety of Aeronautics, Italian Ministery of Defense) for the study of decision processes. I thank heartly Ten.Col. Giuseppe Fauci and Ten.Col. Sandro Monti for the precious collaboration and the insightful and accurate interviews. Those interviews have been conducted by the author at the offices of the Ministry of Defense, Rome, in March–April 2013.

‘unforeseeable events’ are categorized in *types of variables* (categories of events). The broader partition is that between human error and technical variance; and within them various sub-categories are defined. For example, as human factors, categories as ‘situational awareness’, ‘spatial disorientation’, ‘language barriers’. In that way, models of effective corrective action can be constructed on the basis of the observation of occurrences and/or of causal modeling based on pertinent sciences—both natural (physics, engineering, geology, etc.) and human (especially cognitive and social psychology, neuro-sciences); and even ordinal judgments about their probability of occurrence (high to low) can be expressed (cfr [24]). In practice:

- All and every flight is preceded by a briefing, reviewing large check lists of factors, on the basis of available accumulated experience and knowledge.
- All and every flight is followed by a de-briefing, reviewing events, actions and consequences, so that they are memorized and new knowledge produced and transferred.
- An analysis of the causes of occurrences, of the cause-effect relations between the actions that have been found (‘*causal modeling*’) and that could have been found (‘*counter-factual thinking*’)—and the consequences observed are performed; so as to produce valid models of action that can be applied in a fast mode in subsequent occasions.

This process amounts to constructing ‘off-line’ casual models of anomalies and of codifying that knowledge so that it is ready for fast use when ‘unpredictable’ events occur. Paraphrasing Weick’s and colleagues notion of ‘swift trust’ [25], a sort of ‘swift knowledge’ seems to be at the basis of much high reliability action. This class of heuristics offers a way out from a variety of philosophically naive oppositions that populate decision making literature, such as between theory and experience, ex-ante and ex-post rationality, learning by doing and learning by thinking, tacit/non-transferable/experiential knowledge and explicit/transferable/scientific knowledge. The knowledge is both valid *and* embodied into pilots; both tacit and explicit. Not only experiential knowledge is transformed into systematic explicit and tested knowledge; but also, once checked and tested, explicit knowledge is re-transformed into embodied skills so as to liberate attention for new discretionary problem solving. This is what the second class of rational heuristics for fast and reliable action are about.

3.2 Anticipation and C ognition Expansion

During action, rather than shifting attention foreward to ex-post rationality (from foresight to learning by doing), attention is shifted backwards to signals and ‘precursors’ of possible occurrences.

Continuous scanning is achieved through the division of labor between the ‘flying’ and the ‘monitoring’ pilot. They act as an *augmented brain* in at least two ways.

First, thank to the former ‘prevention heuristics’ the ‘automatic pilot’ within the human pilot is expanded, so that attention is liberated for considering new facts and problems.

Second, the two persons are in close *continuous contact*, so that information and action are checked almost continuously and in real time. ‘Gates’ or check points in specific moments of the flight (e.g. once a safety cruise high and speed has been reached, before lending) are pre-defined, in which a check-listed series of controls are to be performed before ‘crossing’ the gate.

Third, *language is highly formalized* into technical terms of unambiguous meaning, so as to avoid as far as possible any misunderstanding. A lot of effort in training is put on ‘listening’ and not discarding any signal that the monitoring pilot may give to the flying pilot. After flight analyses include the improvement and modification of terms identifying procedures, objects and places so as similarities and therefore the likelihood of mistakes are reduced.

3.3 Hypotheses Generation and Testing in Real Time

The still pending core question, though, is how do pilots reason in front of events or problems that have not been previously typified and for which a model of possible effective actions has not yet been constructed.

Surprisingly, the emerging procedures include some of the main rational heuristics identified for logically sound decision making and problem solving under uncertainty in general [25]:

- Making *quick causal diagnoses* (as, say, the ‘old style’ good doctors): a mixture of science and experience allowing to generate many possible diagnoses and prescriptions in a very short time.
- *Using analogy* (drawing possible actions from long series of past observations/experiences). For example, in pilot training, long series of case studies of accidents are proposed through high emotional impact videos, so that the situations and solutions stick into memory and are easily retrievable (a ‘positive’ use of the availability heuristic). Various real disaster avoidance cases and flight simulations show that indeed analogy with past situations is used.
- Ask *whys* and *check data* before taking action; avoid jumping to conclusions and taking unverified action.
- *Ask disconfirming and counter-factual questions*: for example, what if we were to land/turn/do X? Is a ‘short landing’ procedure (touching the lane at its very beginning as it is or is assumed to be short), in principle feasible, compatible with the specific aircraft and weather conditions? Is a planned climbing maneuver, usually within the range of possible performance of an helicopter, actually feasible in conditions of extreme height and temperature?

- Look for actions that are not irreversible, have multiple exits, allow multiple options that may be exercised according to circumstances—a case in *robust action heuristics* [14]—for example fly to a safe zone in which it can be possible to try to fix anomalies while flying, gathering further information from various ground sources, and/or from which emergency landing is easier.

Many of those heuristics are recognizable in the following description of a process of solution of a truly unforeseen anomaly, for which a repertory of solutions, or of procedures for defining solutions, was not included in the flight manuals.

> Not by the books[4]
>
> As many have discovered in the history of aviation, there are sometimes situations where "the book" simply doesn't help. A case in point happened one semi-sunny Arkansas summer afternoon at the Little Rock Air Force Base, Ark., C-130 Center of Excellence "schoolhouse."
>
> The mission for the day was rather routine: Get the mighty C-130J "Super Hercules" airborne, fly a couple of low-level tactical routes, and end the day with some touch-and-go/assault landing practice.
>
> The crew had completed two tactical low-level routes without incident. But the flight was far from over. Cleared inbound on the visual overhead approach, McAlevey called for "gear down."
>
> Snow moved the gear handle to the down position, and that's when the "smooth" mission got rough.
>
> The Advisory, Caution, and Warning System that provides visual and audible indications when malfunctions are detected. Hearing the "caution" sounds through their headsets, the pilots looked down to see "RIGHT GEAR NOT DOWN" on their flight management system displays.
>
> "*The right gear light is not on, eh*," the Canadian co-pilot said.
>
> "*Roger, let's get a place to hold and run the check-list*," McAlevey responded.
>
> The crew contacted air traffic control, and five minutes later found themselves holding at a nearby navigational aid running the "Landing Gear System Failure" check-lists. Among a host of other things, these check-lists require the loadmaster to visually inspect the landing gear assembly from inside the aircraft. That was easier said than done.

[4] Source: *Sicurezza del volo*, *291/2012,* p. 11–19. Case written by Ten. Joshua Fulcher, Little Rock, Arkansas; published on the review "Flight Comment"/2010; translated by T.Col. Massimiliano Macioce.

> The aircraft had a significant load in the cargo compartment that was to be used for ground training after landing. To reach the landing gear access panels in the cargo compartment, Carter and Year had to move the pallets while the plane was in flight no easy task on an airborne aircraft in a holding pattern. Despite the difficulty of the challenge at hand, the loadmasters moved the loads, removed the panels and performed the visual inspection in accordance with the check-list.
>
> Faced with a unique landing gear malfunction, a C-130J crew had to get innovative to return home safely. Once eyeballs were on the affected landing gear, the loadmasters knew they had a significant problem on their hands. Not only had the gear not moved from the up position, Carter noted multiple broken components on the gear itself.
>
> The next 20 minutes were spent following the check-list guidance and trying to get the gear down via alternate methods in the book, but none of them worked. In the process of trying to lower the gear, the crew contacted multiple ground agencies, including Lockheed Martin technical support, which offered suggestions on how to best deal with this emergency. Using an iPhone camera, the crew sent pictures of the structural damage to the maintenance professionals on the ground, which were analyzed and used to help guide them.
>
> Finally, the loadmasters managed to partially lower the gear. But how to fix them? The normal system could not work because the gear was not completely down. The sergeant's extensive experience as a former "E" and "H-model" C-130 loadmaster gave one great advantage in this situation: he had used chains to secure unsafe landing gear in the past. Using chains for landing gear malfunctions is not covered in the J-model flight manual because of a different tie-down mechanism specifically designed for that aircraft. The quick-thinking instructor was able to get chains around both the forward and aft gear assemblies and secure them in place for landing.
>
> All said and done, the aircraft held for two hours, losing fuel weight, prior to the crew making the rather tense final approach and landing. The loadmasters' innovative method to secure the gear, not covered in the J-model flight manual, worked swimmingly, and the gear did not collapse. The fix held and the plane sustained no further damage. The crew shut down the aircraft on the runway and walked safely away from only a minor mishap.

The article concludes urging to remember that "it could have been far worse". For instance, "one of the worst situations a C-130 crew can find themselves in is when the landing gear on one side collapses. Thus, it will impact the ground if one of the main landing gear collapses and the other stays down and locked. The wing extends another 20 ft or so beyond the outboard propeller and will also hit the ground in this situation—which at landing speed means only one thing… disaster." It is then asked "How did the Little Rock Air Force Base, Ark., C130J crew turn a potentially catastrophic emergency, not covered in technical orders, into a relatively routine landing?"

Three factors are outlined that can be reformulated as rational heuristics as follows. First the *knowledge of the causal texture of the technical system and the vast experience of anomalies and solutions*—from which the possibility of using chains for fixing gears was quickly retrieved.

Second, it is noticed that many different ground entities contributed to the crew's success that day (traffic control, maintenance, even the aircraft constructor). "*At this juncture an accurate flow of information is vital.*" Gathering key valid and reliable information is possible, and fundamental, even in a very brief time.

Third, it is noticed that, paradoxically, the problem was solved by ‘breaking the rules’, in the sense of intelligently and informatively amending some normal procedure—in the case the prohibition of using iPhones in flight. In that respect, the commentators invite to also consider one of the first sentences in the flight manual itself, “*…This manual provides the best possible operating instructions under most circumstances, but is a poor substitute for sound judgment. Multiple emergencies, adverse weather, terrain,* etc.*, may require modification of the procedures.*” …“*Is well known that check-lists, unquestionably, cannot anticipate all possible occurrences.*”

The question is how people can revise procedures quickly and soundly. The answer that can be found in the article is that pilots are instructed to always “have doubts and be curious” during his/her training; and instructors to provide “all necessary *explanations* of systems, normal and emergency procedures.” This amount to gaining a *causal understanding* of the working of the system, and of the safety rules themselves, so that in case of failures of check-listed actions, operators are able to intelligently revise check-list items. In fact the article concludes by saying: “Now it’s time for you *to stop and think to find out if there is this any prohibition on your aircraft* (e.g. on cell phone use), *the reasons why* this ban has been applied and *how its use affects the aircrafts systems* (*if any*).”

It may seem strange, but heuristics for causal analysis, hypothesis testing, and knowledge generation—even in interaction among various actors—can be applied in very short times, measurable in minutes. They may well be less wide than rational hypotheses formulation and testing in other settings—such as strategy making or scientific discovery—lasting more extended periods of time, measurable in months or years, but their logical structure does not seem to be qualitatively different.

These results reinforce the idea that we should put into question the common, well established view that there is a plain trade-off between the speed of decision making and its accuracy. Decision processes that are simultaneously fast and correct, are likely to process more rather than less information than processes led by sequential and experiential heuristics—as already noticed by Eisenhardt [9] in a study on firms’ strategic decision making. Some of the effective heuristics for high velocity environments, identified in that study (taking months), are also found in the analysis of pilot decision processes (taking minutes): continuous real time information, considering simultaneously multiple alternatives and possible exits so that action can be re-directed, anticipating rather than waiting for events to occur, consultation among parties with one party entitled to integrate information and to make a final choice.

This last procedure is a ‘social heuristics’ of ‘consultation’. Listening to others is imperative, but residual authority to integrate all inputs and ‘close’ the process is granted to one responsible actor, if velocity is vital. This procedure is in fact adopted not only on airplanes, but also in most other troubled/high velocity environments. A residual right and obligation to close the process in a short time—a rational social heuristics rather than a potentially biasing cognitive short-cut—seems to emerge as the main, or even the only, qualitatively different feature of rational decision making in need to be fast, with respect to the more quiet and long-term conditions of other kinds of problem solving processes, such as research and development.

References

1. Bandura, A.: Social Foundations of Thought and Action. Prentice-Hall, Englewood Cliffs (1986)
2. Bingham, C.B., Eisenhart, K.M.: Rational heuristics: the 'simple rules' that strategists learn from process experience. Strateg. Manag. J. **32**, 1437–1464 (2011)
3. Bonifati, G., Villani, M.: Forth 'exaptation and innovation processes'. In: Grandori, A. (ed.) Handbook of Economic Organization. Integrating Economic and Organization Theory. Cheltenham, UK Edward Elgar (2013)
4. Browne, G.J., Pitts, M.G.: Stopping rule use during information search in design problems. Organ. Behav. Hum. Decis. Process. **95**, 208–224 (2004)
5. Campbell, D.T.: Blind variation and selective retention in creative thought as in other knowledge processes. Psychol. Rev. **67**, 380–400 (1960)
6. Cholakova, M., Grandori, A.: Effective Heuristics for Decisions Under Uncertainty: Lessons from Angel Investing. Academy of Management Meeting, Boston. SSRN abstract_id = 2005154 (2012)
7. Cohen, M.D., Burkhart, R., Dosi, G., Egidi, M., Mareno, L., Warglien, M., Winter, S.: Routines and other recurring action patterns of organizations: contemporary research issues. Ind. Corp. Change **5**(3), 653–698 (1996)
8. Cohen, M.D., March, J.J., Olsen, J.P.: Ambiguity and Choice in Organizations. Universitaetforlaget, Bergen (1976)
9. Eisenhardt, K.M.: Making fast decisions in high velocity environments. Acad. Manag. J. **32**, 543–576 (1989)
10. Felin, T., Zenger, T.: Entrepreneurs as theorists: on the origins of collective beliefs and novel strategies. Strateg. Entrepreneurship J. **3/2**, 127–146 (2009)
11. Fiet, J.O.: The Systematic Search for Entrepreneurial Discoveries. Quorum Books, Westport CT (2002)
12. Gigerenzer, G., Todd, P.M., ABC Research Group.: Simple Heuristics that Make Us Smart. Oxford University Press, Oxford (1999)
13. Grandori, A.: A prescriptive contingency view of organizational decision making. Adm. Sci. Q. **29**, 192–209 (1984)
14. Grandori, A.: A rational heuristic model of economic decision making. Rationality Soc. **22**(4), 477–504 (2010)
15. Grandori, A.: Models of rationality in economic organization: economic, experiential, epistemic. In: Grandori, A. (ed.) Handbook of Economic Organization. Integrating Economic and Organization Theory. Cheltenham, UK: Edward Elgar (2013a)
16. Grandori, A.: Epistemic Economics and Organization. Forms of Rationality and Governance for a Wiser Economy. Routledge, London (2013b)
17. Grandori, A., Cholakova, M.: Unbounding bounded rationality: heuristics as the logic of economic discovery. Int. J. Organ. Theor. Behav. Symp. Models Rationality Decis. Mak. Organ. (2013) (in print)
18. Grandori, G.: Paradigms and falsification in earthquake engineering. Meccanica **26**, 17–21 (1991)
19. Henderson, R., Orsenigo, L., Pisano, G.P.: The pharmaceutical industry and the revolution in molecular biology: Interaction among scientific, institutional, and organizational change. In: Mowery, D.C., Nelson, R.R. (eds.) Sources of Industrial Leadership. Cambridge University Press, Cambridge (1999)
20. Kahneman, D., Slovic, P., Tversky, A. (eds.): Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge (1982)
21. Kiss, O.: Heuristic, methodology or logic of discovery? Perspect. Sci. **14**(3), 302–317 (2006)
22. Lakatos, I.: The history of science and its rational reconstructions. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association VIII: 91–136. Springer, Dordrecht NL (1970)
23. Liedtka, J.: In defense of strategy as design. Calif. Manag. Rev. **42**(3), 8–30 (2000)

24. Lodato, G.: ‘L’Operational Risk Management a livello operativo: un case study’ Sicurezza del Volo. **283**, 16–27 (2011)
25. Magnani, L., Nersessian, N.J., Thagard, P. (eds.): Model-Based Reasoning in Scientific Discovery. Kluwer Academic Publishers, Dordrecht (1999)
26. Meyerson, D., Weick, K.E., Kramer, R.M.: Swift trust and temporary teams. In: Kramer, R.M., Tyler, T.R. (eds.) Trust in Organizations, pp. 166–195. Sage Publications, Thousand Oaks (1996)
27. Payne, J.W.: Task complexity and contingent processing in decision making: a replication and protocol analysis. Organ. Behav. Hum. Perform. **16**, 366–387 (1976)
28. Popper, K.R.: The critical approach versus the mystique of leadership. Hum. Syst. Manage. **8**, 259–265 (1989)
29. Rasmussen, J.: Risk management in a dynamic society: a modeling problem. Saf. Sci. **27**(2–3), 183–213 (1997)
30. Sarasvathy, S.: Effectuation: elements of entrepreneurial expertise. Edward Elgar, Cheltenham, UK (2008)
31. Savage, L.J.: The Foundations of Statistics. Wiley, New York (1954)
32. Simon, H.A.: A behavioral model of rational choice. Q. J. Econ. **69**, 99–118 (1955)
33. Simon, H.A.: From substantive to procedural rationality. In: Latsis, S.J. (ed.) Method and Appraisal in Economics. Cambridge University Press, Cambridge (1976)
34. Simon, H.: Models of Discovery and Other Topics in the Method of Science. Reidel, Dordrecht and Boston (1977)
35. Sjöberg, L.: Rational risk perception: utopia or dystopia? J. Risk Res. **9**(6), 683–696 (2006)
36. Schwing, R.C., Albers Jr, W.A. (eds.): Societal Risk Assessment: How Safe is Safe Enough?. Plenum Press, New York (1980)
37. Weick, K.E., Sutcliff, K.M.: Managing the Unexpected. Wiley, New York (2007)

Dynamic Generation of Hypotheses: Mandelbrot, Soros and Far-from-Equilibrium

Emiliano Ippoliti

Abstract In this paper I argue that the effective way to account for the behavior of stock market prices is to put in use a dynamic approach—bottom-up, local, non-axiomatic, heuristic. To this end, I provide an analysis of the generation of four main hypotheses used to explain stock market prices (SMP). In particular I show how the means of generating these hypotheses is essential to assessing their efficiency and plausibility. In formulating a hypothesis, a selection of features of SMP is made for incorporation in a theory. This selection may be expressed mathematically in most of the cases. An examination of these means of generation can show us why some of these hypotheses are successful and efficient and some not, and can also shed light on the extent to which a particular hypothesis can be usefully applied. Thus the study of the means of generation of hypotheses will offer us a guide to formulating new hypotheses in a reliable and cogent fashion.

1 Introduction

In this paper I argue that the way to effectively account for the behavior of stock market prices (SMP) is to put in use a dynamic approach—bottom-up, local, non-axiomatic, heuristic (see e.g. [1, 2]). To demonstrate this I provide an analysis of the generation of four main hypotheses used to explain SMP.

In particular I examine the construction of one cogent hypothesis for SMP, that is the far-from-equilibrium hypothesis. To this end I analyze the generation of the hypotheses that preceded the far-from-equilibrium hypothesis.

(1) The Efficient Market Hypothesis (EMH), pointing at its main vulnerability, the idea of 'equilibrium', which does not enable us to explain booms-and-busts, or at least their frequency.

E. Ippoliti (✉)
Sapienza University of Rome, Rome, Italy
e-mail: emiliano.ippoliti@uniroma1.it

E. Ippoliti (ed.), *Heuristic Reasoning*, Studies in Applied Philosophy, Epistemology and Rational Ethics 16, DOI 10.1007/978-3-319-09159-4_8

(2) The Fractal Market Hypothesis (FMH), which offers a new interpretation of the data, shows new properties of financial markets, and undermines the effectiveness of the notion of equilibrium. Even though it does not explain the reasons for these properties and does not offer predictions that can be put to use—due to the sensitivity to initial conditions—it generates new mathematics and explain to us when we can expect markets to be stable.
(3) The Reflexive Market Hypothesis developed by George Soros (RMH), which has received little scholarly attention but offers a cogent qualitative explanation of several properties identified but not explained by (2). This hypothesis draws on the distinction between endogenous and exogenous forces in the behaviour of prices and it enables us to explain boom-and-bust and crashes.

The Far-form-Equilibrium Hypothesis (FEH) relies on the distinction between exogenous and endogenous forces and sets out to develop a means to forecast crashes and bubbles, also the flash-crashes (e.g. [3, 4]).

The main point of this paper is to show, using as an example the SMP, how the means of generating these hypotheses is essential to assessing their efficiency and plausibility. In formulating a hypothesis, a selection of features of SMP is made for incorporation in a theory. This selection may be expressed mathematically in most of the cases. An examination of these means of generation can show us why some of these hypotheses are successful and efficient and some not, and can also shed light on the extent to which a particular hypothesis can be usefully applied. Thus the study of the means of generation of hypotheses will offer us a guide to formulating new hypotheses in a reliable and cogent fashion.

More specifically I will argue that the generation of a new hypothesis has to draw on a preliminary verbal conceptualization (a discourse) on a specific subject, that is a verbal and non-formal description of it, which establishes the entities to investigate, their properties and relations, and a set of variables that affect them. This is a bottom-up process and it is the only way to incorporate the (domain) specific features of the subject in a plausible representation of it, which can possibly end up in a mathematical theory (like the FEH). Thus my thesis is that generation of new hypotheses and, possibly, new mathematics (formal treatments) stem from a preliminary verbal modeling and conceptualization, which delimitate the variables and features of a phenomenon to investigate.

In particular I will show how this four hypotheses draw on different processes of generation: the EMH is generated in a static way—top-down, axiomatically, trying so save a given framework and its formal tools—, while the FMH, RMH and FEH are generated in a dynamic way—bottom-up, locally, trying to reflect and incorporate in verbal or formal model new features. The FHM reflects unexpected statistical features in a new mathematical model. The RMH incorporates in a verbal, qualitative model a critical characteristic, i.e. investors' expectations and the consequent reflexive and endogenous processes. The FEH refines the RMH by modeling herding with bifurcation theory and sets out to offer a way to forecast critical events in SMP.

The generation of hypothesis is the central theme in the pursuit of a 'logic' of discovery, that is a non-mechanical, fallible but rational means to amplify our knowledge. The development of a means to discover requires a different ‘logic’ than the deductive one, but this does not mean that the deductive reasoning has to be put aside—for it is still useful. As a matter of fact, even though deductive reasoning, *strictu sensu*, cannot generate new knowledge and does not offer a means for producing hypotheses (as the conclusion is contained in the premises of its closed set of rules of inference), it is a means to test and control our hypotheses by confronting the compatibility of their consequences with known facts and results.

On the other side, a ‘logic’ of discovery is expected to specify procedures for generating hypotheses starting from a set of data. These procedures and the rules of inference (e.g. analogy, induction, metaphor) are an open set, as they can be extended with new discoveries. Moreover these procedure and rules draws on a preliminary verbal reasoning and conceptualization of the issue.

We need both these ‘logics’ for achieving new knowledge: one logic generates new hypotheses and extends the content of our knowledge. These hypotheses are partially justified and amplify the content of the premises from which they are drawn. The second logic tests and controls the hypotheses by confronting the compatibility of their consequences with known facts.

In particular the SMP problem shows how a preliminary conceptualization and the use of ampliative inference are essential to generate better and better hypotheses and comprehension about a given issue.

2 Prices Trajectories

The explanation and forecast of the SMP is a very interesting problem both from a quantitative and a qualitative view point.

On the former, the SMP is a data-fitting problem and, as such, it is underdeterminated: it admits infinite solutions—an infinite amount of curves (equations) that approximates the data and forecasts the next point of the series. As known, adding new points does not resolve the issue, since the new series of points cuts off whole classes of fitting-curves, but in turn it can be still approximated by an infinity of new curves.

On the latter, the problem requires determining the variables that affect the price of a share in a financial market. It is a hard problem, since two opposite outcomes can be trigged by the same event: “the precise market mechanism that links news to price, cause to effect, is mysterious and seems inconsistent. Threat of war: Dollar falls. Threat of war: Dollar rises. Which of the two will actually happen? After the fact, it seems obvious; in hindsight, fundamental analysis can be reconstituted and is always brilliant. But before the fact, both outcomes may seem equally likely. So how can one base an investment strategy and a risk profile entirely on this one dubious principle: I can know more than anybody else?” [5, p. 8]. The problem is that these variables are virtually infinite and prices are the result of the actions of investors who sell and buy them: the individual and collective decisions (i.e. behavior) of investors cause the price to move up and down.[1] The balance between supply and demand set the price of an asset: the price moves up if there are more buy than sell orders and it moves down vice versa. Obviously the problem here is that stock markets future is essentially uncertain (not simply risky) and therefore the traders' evaluations of companies' future profitability will change. It is difficult, to say the least, to foresee future directions of SMP even for time scales of the order of decades, for which one could hope for a negligible influence of 'noise' (see e.g. [6, 7]).

Several hypotheses have been put forward to approach the problem. Accordingly it is an interesting case study of the ways of generating hypotheses. These hypotheses can be broken down in two main classes, the random walk and non-random walk.

The former expresses the neoclassical view, which draws on the Efficient Markets Hypothesis (EMH) and its three versions—weak, semi-strong and strong. These hypotheses share the idea that prices have a random behavior, which is determined by changes in information about future prospects arriving in a random and not predictable fashion.

The latter, by questioning and criticizing the core of neoclassical approach, contrasts this view and argues that prices trajectories are not random at all and exhibit patterns. In this paper I will consider only a subset of this class[2]: the Fractal Markets Hypothesis (FMH), the Reflexive Markets Hypothesis (RMH) and the Far-from-Equilibrium Hypothesis (FEH). These hypotheses can be seen as parts of the construction process of the same approach to the problem. I will show how the means and inferential processes employed to generate them are essential to explain why some of these hypotheses are successful and effective and some not, why some of them select and encapsulate specific features of SMP, and can also shed light on the extent to which a particular hypothesis can be usefully applied. Thus the study of the means of generation of hypotheses will offer us a guide to formulating new hypotheses in a reliable and cogent fashion.

To this end, I will examine how these hypotheses conceptualize and model one of the most puzzling problems in SMP—that is bubbles (boom-and-bust) and crashes. A speculative bubble generates a price trajectory such that a constant upward trend starting from a point in time t is suddenly reversed, pushing the prices down towards the same level at point t in a relatively short period. An example of bubbles (see Fig. 1) is the 1998–2000 Dot.com bubble.

---

[1] This behavior is based on two simple rules that guide investment decisions:

(a) if the investor expects that the market is going up in the future, he should buy and hold the stock until a change of direction of the price occurs (i.e. to be 'long' in the market);

(b) if the investor expects that the market is going down, he should do nothing, sell if he can by borrowing a stock and giving it back later by buying it at a smaller price in the future (i.e. to be 'short' in the market).

[2] For instance I will not consider the Inefficient Markets Hypothesis (IMH) proposed by Haugen [8] and the Financial Instability Hypothesis (FIH) put forward by Minsky [9].

Fig. 1 Dot.com (Internet) bubble

A crash is a sudden and sharp drop of an asset's value in a small interval of time (normally few days). A recent example (see Fig. 2) is the gold crash in April 2013, which took the price of gold down by more than 16 % in five days.

A received view on these events is the one provided by Galbraith [10] and Kindleberger [11], according to which the key factors causing bubbles and crashes are easy credit creation—"that is why so many bubbles have their origins in the sins of omission or commission of central banks" [12, p. 119]—and an 'irrational' euphoria that pushes the public (but not all the traders) to invest in a bull market. More specifically, Galbraith argues that:

(a) the euphoria pushes individuals and institutions to believe in a better future, i.e. that they are going to be richer, and to ignore or explicitly dismiss what conflicts with this belief.
(b) An expansion of credit in the form of brokers' loans pushes the investors to buy stocks and become dangerously leveraged.[3]

In addition, Kindleberger maintains that a bubble typically involves five steps:

1. Displacement. The emergence of an opportunity[4] pushes the market up.
2. Euphoria. This opportunity generates euphoria of rising prices, while an expansion of credit feeds the bubble.

[3] Galbraith examines also the role played by asymmetric information and cross-border capital flows. In the former case, insiders—i.e. those concerned with the management of bubble companies—know much more than the outsiders and in a bubble exploit this asymmetry fraudulently. In the latter case, bubbles are more likely to occur when capital flows freely from country to country.

[4] For instance: a new economy with novel colonial possibilities, new markets, an increasing currency, new technologies, some dramatic political change.

Fig. 2 Gold crash in April 2013

3. Mania. The euphoria becomes a mania: people invest into stocks, commodities, real estate, confident to become rich without a clear understanding of what is going on.
4. Distress. The markets stop rising and people who have borrowed heavily find themselves overstretched. Unexpected failures occur.
5. Revulsion. In the final phase a self-feeding panic bursts the bubble. People dump in a hurry whatever they have bought at greater and greater losses and cash rules.

Of course this description is incomplete: for instance it does not account for the role of 'fundamentals' in a bull market or for the factors triggering the speculative mania. Galbraith specifies only weak qualitative 'marks' of the emergence of the bubble[5] and, in the end, the received view is vague about the causes of a market crash, and simply argues that almost any event could have pushed irrational investors to sell towards the end of bubble, not really explaining the reasons for the crash.[6]

---

[5] E.g.: margin buying, the formation of closed-end investment trusts, the transformation of financiers into celebrities.

[6] More refined explanations have introduced other potential factors as crucial in the production of crashes and bubbles: computer trading, derivative securities, illiquidity, trade and budget deficits, overvaluation, the auction system, the presence or absence of limits on price movements, regulated margin requirements, off-market and off-hours trading (continuous auction and automated quotations), the presence or absence of floor brokers who conduct trades but are not permitted to invest on their own account, the extent of trading in the cash market versus the forward market, the identity of traders (i.e., institutions such as banks or specialized trading firms), the significance of transaction taxes (see [7], pp. 5–6).

3 The Neoclassical View: The Efficient Market Hypothesis

The neoclassical approach to the SMP is condensed in the Efficient Markets Hypothesis (EMH) (see [13–15]) and its refinements, which offer an answer to both the qualitative and the quantitative side of the issue.

As regards the qualitative side, the EMH states that investors take decisions aiming at maximizing their utility on the basis of expected returns and risk and that the prices will be fixed by the equilibrium point between demand and supply. The stock market's volatility, the prices' variations, is the outcome of the random, external arrival of new information that affects the equilibrium's value of shares: no arrival of new information from outside the market, no reaction by the investors, no changes in prices. Moreover, when new information arrives, the EMH states that investors determine the value of capital assets in a very specific and rational way: "they objectively consider information about the investment opportunities offered by different companies, and data about world economic prospects. Information that affects the future prospects of investments arrives randomly, generating random movements in the expected future prospects of firms. Investors' rational appraisal of this information leads to an efficient valuation of shares on the basis of expected return and risk, with price variations being caused by the random arrival of new information pertinent to share prices" [16, p. 234].

The EMH sees efficiency[7] as the movements of financial prices as an immediate and unbiased reflection of incoming news about future earning prospects: "the deviations from the random walk observed empirically would simply reflect similar deviations in extraneous signals feeding the market" [7, p. 133]. In particular this implies that an efficient market is such that an asset price equals its 'fundamental value', i.e. the discounted sum of expected future cash flows where, in forming expectations, the investors correctly process all available information. In turn, this implies that in an efficient market there is "no free lunch": no investment strategy can earn excess risk-adjusted average returns, or average returns greater than are warranted for its risk. So EMH establishes that rational speculative activity would not only eliminate

[7] Keen points out that efficiency in EMH can mean "at least four things: that the collective expectations of stock market investors are accurate predictions of the future prospects of companies; that share prices fully reflect all information pertinent to the future prospects of traded companies; that changes in share prices are entirely due to changes in information relevant to future prospects, where that information arrives in an unpredictable and random fashion; and that therefore stock prices 'follow a random walk,' so that past movements in prices give no information about what future movements will be—just as past rolls of dice can't be used to predict what the next roll will be" ([16], pp. 244–245).

riskless arbitrage opportunities, but also that the presence of many sophisticated traders in the market will burst a bubble at the very beginning.[8]

One of the strength points of the EMH is that the explanation is based on decisions and actions of investors. The random price trajectories are achieved through the active participation of many investors seeking greater wealth: this herd of investors actively analyzes all the available information and takes investment decisions based on them. Accordingly, as Bachelier and Samuelson argued, "any advantageous information that may lead to a profit opportunity is quickly eliminated by the feedback that their action has on the price. Thus, the price variations in time are not independent of the actions of the traders, but derive from them. If such feedback action occurs instantaneously—in an ideal world of ideal "frictionless" markets and costless trading—then prices must always fully reflect all available information and no profits can be garnered from information-based trading (because such profits have already been captured)" [7, p. 40–41].

One of the main consequence of the EMH is that market cannot be beaten: "in a perfect capital market, price fluctuations simply reflect the random arrival of new information, and yesterday's price trends are as relevant to tomorrow's as the last roll of the dice is to the next" [16, p. 255]. So we have that "the more active and efficient the market, the more intelligent and hard working the investors; as a consequence the more random is the sequence of price changes generated by such a market. The most efficient market of all is one in which price changes are completely random and unpredictable" [7, p. 41].

In order to be obtained this characterization of SMP (equilibrium, no free lunch, unpredictable and random markets, no bubbles) the EMH has to assume no limited arbitrage and the hypothesis that investor have identical, accurate expectations of the future, equal access to unlimited credit, and have the capability of updating their beliefs correctly when they receive new information, so that the subjective distribution they use to forecast future determinations of asset prices and returns is indeed the distribution from which those determinations are drawn.

On the other hand, as regards the answer to the quantitative side of the SMP, the EMH expresses the alleged randomness of prices trajectories with the Brownian Motion, that is conjecturing that the prices can be plausibly approximated (see [17]) by the differential equation $\frac{\partial S}{S} = \mu dt + \sigma dW$, where W is Wiener process with mean equal to zero and variance dt. This differential equation shapes the most important models of stock prices based on the EMH, like the Merton-Scholes-Black equation.

In effect, the EMH is supported by sets of data that, at first sight, seem to make it plausible and reliable: some features of the available data about SMP seem to

[8] A typical example is the following: suppose that the fundamental value of a share of Yahoo! Is \$30. Imagine that a group of irrational traders becomes excessively pessimistic about Yahoo!'s future prospects and through its selling, pushes the price to \$20. Defenders of the EMH argue that rational traders, sensing an attractive opportunity, will buy the security at its bargain price and at the same time, hedge their bet by shorting a "substitute" security, such as Google, that has similar cash flows to Yahoo! in future states of the world. The buying pressure on Yahoo! Shares will then bring their price back to fundamental value of \$30 [4, pp. 3–4].

support the EMH (see e.g. [7, p. 34–36]), like the distributions of daily returns and the correlation function of the returns at the minute time scale. In effect:

(a) The distributions of daily returns (e.g. DJIA and the Nasdaq index for the period January 2, 1990 until September 29, 2000), show that positive and negative returns are almost identical, so that there is almost the same probability for a price to increase or decrease. They appear to be random.
(b) The correlation function of the returns at the minute time scale (e.g. the Standard and Poors 500 futures for a single day, June 20, 1995), shows that returns are essentially uncorrelated beyond a few minutes in active and well-organized markets: “as a consequence, successive returns cannot be predicted by linear extrapolations of the past” [7, p. 36].

Secondly, EMH is supported by the observation that on average and in the long run even the smartest investors find it hard to do better than the comprehensive common-stock averages, such as the Standard and Poors 500, or even better than a random selection among fundamentals of financial markets stocks of comparable variability.

Unfortunately the data tell with the same strength also a different story. First of all, it is possible to continually beat the market. A classic example is the performances of Soros’ Quantum Fund, which earned an average of 32 % annual return from 1969 to 2000: if you had invested €1,000 in 1969, by reinvesting the whole capitalized amount you would have €4 million in 2000. Moreover the EMH was formulated on the basis of a (comparatively) small set of data (1950–1964 prices) and above all the EMH draws on a narrow idea of the investors’ decisions and behaviour, reflected in the assumption that all investors agree about the valuations of all companies. This assumption is crucial to obtain equilibrium: it is this convergence of expectations and valuations to ‘produce’ the equilibrium. Furthermore, this assumption cuts off the effect of potential feedback between valuations and perceptions, or better EMH assumes that is a one-way process that goes from the first to the second, and de facto it states that investors can be uninterested in the behavior of the other investors (since an investor knows that the other thinks exactly what he does). In effect, if investors are allowed to disagree about the future prospects of companies, the future won’t be always as the investors expect. This divergence between expectations and outcomes will generate disequilibrium in the stock market. Moreover, if investors influence each other’s expectations, then the market can be dominated by pessimistic and optimistic mood, and the market will have cycles fed by the alternation of one dominant mood to the other.

Sure, the EMH could be defended by arguing “that the trends we can observe in commodity and financial markets are merely temporary aberrations which will be eliminated in the long run by the ‘fundamental’ forces of supply and demand” [18, p. 21], but this requires the controversial assumption that investors’ expectations and fundamental forces have to be independent.

In addition empirical economics has found plenty of facts that are incompatible with the EMH’s predictions and leading to inefficiencies (see e.g. [7, p. 87–88]). First of all, it is well-recognized that prices’ movements are too large compared with the EMH’ predictions (even considering the costs of gathering information).

Furthermore, the realization of efficient markets can be prevented by trading rules, since inadequate methods of pricing may lead to a slow and inefficient convergence to the equilibrium price or even produce a divergence from it. Again, it is controversial that a public announcement will produce common expectations among the traders, since each trader can still be uncertain about how others will use this information. Furthermore, the phenomenon of "herding" can also be considered an example of market failure, as it leads to important deviations from "fundamental" or "equilibrium" prices.

All these imperfections and inefficiencies have generated refinements of the EHM—the weak, the semi-strong, and the strong version.[9] But also with these improvements, the EMH shows strong flaws, in particular the lack of an explanation (not to say a forecast) of the emergence and frequency of crashes and boom-and-bust. In effect, "from an efficient market viewpoint, the speculative attacks are nothing but the revelation of the instability and the means by which markets are forced back to a more stable state" [7, p. 250]. Moreover, according to the EMH, large price moves like bubbles and crashes should only occur with high-impact news.

The point of my paper is to show how these flaws are the result of the specific process of generation of the EMH, in particular of the methods, inferential means and models employed in its formulation. In particular I argue that this process is static, that is aiming at integrating (part of) the data in a top-down, axiomatic fashion, looking for 'rigor' and compatibility with pre-existing theories rather than for genuine problem-solving and cogent interpretation of the data.

[9] The three versions of the EFM basically differ on the way an information about a company is supposed to spread. The weak-form states that since even an updated public info does not to spread easily, freely and quickly, the people who know it can profit by beating the people who does not know it yet. Hence some kinds of fundamental analysis may provide returns, since it can help to predict prices' changes. But technical analysis cannot be effective, since past prices' changes is common public knowledge. Future prices cannot be predicted by analyzing prices from the past: returns cannot be made *in the long run* by using investment strategies based on historical share prices or other historical data—prices have no serial dependencies. This implies that future price are determined by information not contained in the price series. Hence, prices must follow a random walk. A major point here is that this form of EMH does not imply that the prices remain at (or near) equilibrium, but simply that investors won't be able to *systematically* profit from market 'inefficiencies'. The semi-strong-form states that information flows very quickly so that public information is useless and only inside information can be trade on to get a return. The inside information gives a slight advantage that can be traded and profited. Accordingly the semi-strong-form efficiency implies that neither fundamental analysis nor technical analysis will be able generate excess returns, since they are based on public information. The strong-form states that information flows freely and instantaneously so that prices reflect all the relevant information, public and private, and no one can earn excess returns. This means that information in general is useless: inside information, fundamental or technical analysis are not means to predict the futures' prices. There is no way to beat the market, which is completely random. Note that even the fact that some money managers continually beat the market does not refute the strong-form efficiency: few great performers are expected, for the existence of hundreds of thousands of fund managers worldwide imply a *normal* distribution of returns.

In effect, the EMH draws on an 'opportunistic' selection of the data about SMP: the part of the data that shows the features that can be treated in terms of equilibrium and randomness, such as the distributions of daily returns and the correlation function of the returns at the minute time scale. This selection allows the transfer to SMP of some properties and features from physics (the so-called 'proto-energetic' physics, see [19][10]). Tellingly, these features are the right ones for a 'neoclassical' interpretation, conceptualization and mathematization. At first sight the concept of equilibrium seems very useful: "it allows us to focus on the final outcome rather than on the process that leads up to it. But the concept is also very deceptive. It has the aura of something empirical: since the adjustment process is supposed to lead to an equilibrium, an equilibrium position seems somehow implicit in our observations" [18, p. 27]. Moreover, equilibrium points can be obtained by simple principles (axioms) concerning few concepts like rationality, utility, demand and supply. "Equilibrium is the product of an axiomatic system. Economic theory is constructed like logic or mathematics: it is based on certain postulates and all of its conclusions are derived from them by logical manipulation" [18, p. 27]. The problem here is that if equilibrium is never achieved, it does not "invalidate the logical construction, but when a hypothetical equilibrium is presented as a model of reality a significant distortion is introduced. If we lived in a world in which the angles of a triangle did not add up to 180°, Euclidean geometry would constitute such a misleading model" [18, p. 27].

A partial defense of the EMH under this respect is the fact that it was formulated before the concept of chaos was constructed and refined, so that it appeared legitimate to focus on states of 'equilibrium'. This choice was also convenient because the economists working in finance theory could use all the mathematical and statistical tools built for random processes, like the ones for Brownian motion. In effect, in this sense, the crucial passage in the construction of the EMH is the analogy between price trajectories and Brownian motion (GBM). Price variations seem to exhibit the casual behaviour that is typical of Brownian motion, so that is seemed reasonable to assimilate the two phenomena. But the similarities employed to propose GMB as a legitimate candidate to approximate price trajectories are really meager and weak (see [17]). Moreover, the integration of the selected data is conducted in a top-down fashion: the starting point is the neoclassical framework, which is thus given, and the mathematical model follows from it. Furthermore, the process of generation of hypothesis 'starts with algebra' (the mathematical concepts and tools—just like GBM—are already known and given)

[10] Mirowski (see [19, p. 63]) states that proto-energetics physics is a set of historical analytical tools that includes the law of conservation of energy and the bulk of rational mechanics, but excludes the entropy concept and most post-1860 developments in physics (i.e. it includes the formalisms of vector fields, but excludes Maxwell's equations, or even Kelvin's mechanical models of light). The term proto-energetics expresses the fact that it resembles the content of the energetics movement. Classical thermodynamics diverges from proto-energetics in one crucial feature: thermodynamic processes only change in one direction. In proto-energetics, time is isotropic, which means that no physical laws would be violated if the system ran backward or forward in time.

and it is based on the search of 'rigor', the compatibility or, better, the derivability of the known mathematical model from the economic orthodoxy, focused on the determination of points of equilibrium.

So the generation of the EMH is flawed at least in two senses:

(a) the EMH can be successful, if possible, only when the features of data about price trajectories fit the ones of a stochastic process, like Brownian motion. But this step is equivalent to the conjecture that SMP have no specific features or, at least, no features different from the stochastic ones.
(b) the generation of the EMH it is based on an ad hoc move, i.e. the accommodation of some features of the data with the neoclassical view. In fact the data are 'sacrificed' to hypotheses, in the sense that a part of the data (the large price variations and their frequency) is deliberately ignored and treated as noise in order to keep the rest of the data compatible with the neoclassical framework. Thus, the process of generation of hypothesis is carried out in a way that seeks for the best available mathematics that approximates the hypotheses, not the best hypothesis for the all the data and their features. The data are interpreted as to conveniently fit the neoclassical hypothesis. Hence, EMH does not produce a solution to the problem, and it is simply an attempt to use a known 'solution' for the problem under investigation: in the end, the problem and the data are modified in order to be treatable by some known and convenient 'solution'. The incessant struggle of the EMH with the data and the continual adjustments made to keep it compatible with them, suggest that the theory seems to commit the fallacy of suppressed evidence, rather than providing a cogent interpretation of the data and their relations. Thus, the 'static' generation of the EMH is at the core of its failure: it cuts off part of data (fat tails, anomalous price variations, etc.) and strategically reduces financial data to the principles of neoclassical economics. As a consequence, as the Mandlebrot's quantitative interpretation of SMP and Soros' qualitative show, the neoclassical, 'proto-energetics' view on SMP is hardly tenable.

4 The Fractal Markets Hypothesis

The generation of Fractal Market Hypothesis (FMH, see [5]) is completely different from the EMH. First of all it is based on a different selection of data about SMP, i.e. the changes in cotton prices on a time series 1900–1960, analyzed by Benoit Mandelbrot [20] who shows that this set of data has two features that do not fit the EMH and cannot be reduced to random processes:

(a) it does not follow a normal distribution[11];

[11] For instance, the standard deviation of daily movements on the Dow Jones is about 1 %. If stock market prices were generated by a normal process, then extreme movements—say superior to 5 % in just one day—would occur 1 in a 1,000,000.

(b) it shows a peculiar self-similar 'pattern'—indifferent to scale: the curve described by price changes for a single day is similar to a month's curve. Moreover, these patterns occur during a period that includes Great Depression and two world wars. Thus they are hardly random.

On the basis of this findings, the FMH provides a new answer to the quantitative side of the SMP problem. (a) and (b) show that bubbles and crashes are, in a discontinuous, dependent, concentrate and scaling fashion, legitimately possible outcomes of the SMP dynamics, and that they are not rare and extreme events. The starting point of the generation of FMH is a statistical, domain-specific feature, which is incorporated in a mathematical theory that shows how these trajectories can be approximated. In effect, price trajectories exhibit specific features that cannot be treated by neoclassical formalisms and models: the FMH incorporates these domain-specific features in a new mathematical model. It solves the SMP problem bottom-up, starting from an interpretation of the specific properties of the data, ending up with the anew account of it. The hypothesis is that stock market prices follow a complex pattern called 'a fractal' and, accordingly, the statistical tools used by the EMH, designed to model random processes, will provide systematically misleading predictions about SMP. Nay, fractal are a better approximation of price trajectories.

In a nutshell, the basic idea behind a fractal is that each number in the series is a simple but nonlinear function of previous numbers in the series. This implies that the next number, the next point of the trajectory, is not independent of all previous numbers. It is not like tossing a coin. This means that the FMH draws on the idea that in stock markets it is quite possible that each price movement is a complex function of previous price movements. But this does not imply that stock markets can be predicated and profited.[12] On the contrary, they cannot predicted by means of fractal models.

The idea that prices are non-linear functions of previous prices opens up the possibility that past movements in prices cause or affect the next ones. That is, according to the FMH is it plausible that the market could be driven by endogenous processes in which just previous price determine future price—but it is almost impossible to predict the way the market will move, and by how much. This implies that the EMH explanation of price trajectories as caused by exogenous, external processes, such as the random arrival of news and information, is at least partial, if not misleading. Thus, according to FMH, highly unstable dynamic systems generate stock prices which appear random, but behind which lies deterministic patterns.

[12] The main reason for that is the property of 'sensitive dependence on initial conditions.' Even if you knew precisely the 'system' which generated the Dow Jones Industrial Average, you could never know the precise value of the index because of rounding error. If e.g. your initial measure of its value is out by 1/10th of a percent—rather than being, say, 10,456.4, it was actually 10,456.5, one day (or iteration) later, your model would be wrong by (say) 1 %; one day later by 10 %; and a day after that, it would be completely useless as a means of predicting the following day's value. This is because any measurement errors you make in specifying the initial conditions of a fractal model grow exponentially with time, whereas for a random model the errors normally grow linearly (and can even fall with time for a stable system).

The fractal objects offers a new statistical way to interpret the data, so that some specific features of the SMP can be expressed and incorporated in this model. Thus, in order to treat and model this features, Mandelbrot built a new mathematics, which approximated the data in a much more cogent and plausible way.

Nevertheless, the FMH does not offer an answer to the qualitative side of the problem, as it can't provide an explanation of these features and trajectories in terms of the behaviour and decisions of the investors. In essence, the FMH is a critique to the idea that price movements in the stock market are random: it is a way to characterize the statistical properties of the market, and not a theory about how the market actually behaves. Nevertheless, as noted by Peters (see [20]) and Keen (see [16]) the FMH offers an explanation of the conditions of stability of the markets, which are in contradiction with the EMH. The market will be stable when "it allows investors with different time horizons to trade smoothly. As a result, heterogeneity—the fact that all investors are not the same—is a vital part of this theory" [16, p. 341].[13] The FMH, therefore, explains the stability of the market starting from a more plausible assumption, that is that investors differ in their time horizons. This tells us also when instability is likely to be generated: it happens when all investors suddenly switch to the same time horizon.

The FMH is generated in dynamic way, that is bottom-up, without an opportunistic selection of the data, and aims at providing a plausible and local interpretation of the data, not an interpretation compatible with a given theory and its general principles. It is the result of the mathematization of new features of SMP. No surprise, then, "that it is more consistent with stock market data, more robust, and completely untainted by any assumption that the market is in, or tends toward, equilibrium" [16, p. 355]. Therefore, in the light of FMH, bubbles and crashes can occur quite often, since this hypothesis is built on the observation that bubbles and crashes are legitimately possible outcomes of the dynamics of stock market price, and not rare and extreme events. Nevertheless, the FMH doesn't offer an explanation about what produces the data and their domain-specific features, it is no able to produce a qualitative interpretation of the data. Soros' Reflexive Market Hypothesis helps us in finding an answer to this issue.

[13] "Take a typical day trader who has an investment horizon of five minutes and is currently long in the market. The average five-minute price change in 1992 was—0.000284 % [it was a 'bear' market], with a standard deviation of 0.05976 %. If, for technical reasons, a six standard deviation drop occurred for a 5 m horizon, or 0.359 %, our day trader could be wiped out if the fall continued. However, an institutional investor—a pension fund, for example—with a weekly trading horizon, would probably consider that drop a buying opportunity because weekly returns over the past ten years have averaged 0.22 % with a standard deviation of 2.37 %. In addition, the technical drop has not changed the outlook of the weekly trader, who looks at either longer technical or fundamental information. Thus the day trader's six-sigma [standard deviation] event is a 0.15-sigma event to the weekly trader, or no big deal. The weekly trader steps in, buys, and creates liquidity. This liquidity in turn stabilizes the market" [21, p. 77].

5 Reflexive Markets Hypothesis

George Soros' account of stock market behaviour provides an answer to the qualitative side of SMP's behavior which is connected with the FMH's main features. On the other side, he does not provide an answer to the quantitative side of the problem.

Soros draws on two ideas: a straight criticism of the EMH hypothesis, inspired by Sraffa's approach (see [22]), and the findings that prices can be affected by 'endogenous forces', i.e. simply by the movements of previous prices. The fundamental idea is that markets exhibit continually, but not continuously, 'reflexive dynamics', that is a double feedback mechanism between the investor's expectations and price behaviour, "which can be observed and converted into profit" [18, p. 21]. In particular, the Reflexive Market Hypothesis (RMH) draws on a large verbal conceptualization and modeling of SMP, on a qualitative analysis that offer also en explanation of the features modeled by the FMH. The explanation, which is based on a fine examination of the investors' behaviour, is much more plausible of the one provided by EMH and even though RHM has received relatively little scholarly attention, it offers a penetrating view on the SMP. The RMH conceptualizes and models the role of investor's expectations in the price's formation and behavior. One of the specific features of social science and especially finance is that "as G. Soros has pointed out, market players are 'actors observing their own deeds'" [7, p. 136]. As stressed by Keen, who follows Keynes' argument, the essence of stock market game:

is not to work out what particular shares are likely to be worth, but to work out what the majority of other players are likely to think the market will think they are worth, since it is not sensible to pay 25 for an investment of which you believe the prospective yield to justify a value of 30, if you also believe that the market will value it at 20 three months hence. In one of the most evocative analogies ever used by an economist, Keynes compared investing in shares to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view [16, p. 292].

Keynes states that, rather than processing rationally investment prospects and world economic conditions, the investors look at each other in order to forecast how the majority will value particular companies in the immediate future. Sure, investors take in consideration world economic conditions, but the main investors analyze and processes info about the investment community itself inside and not outside it. This account for investors' behaviour is radically different from the one described by the EMH. Soros stresses that the EMH conjectures that "markets are always right—that is, market prices tend to discount future developments accurately even when it is unclear what those developments are" [18, p. 14]. But his analysis shows that is much more plausible to accept the opposite point of view:

> I believe that market prices are always wrong in the sense that they present a biased view of the future. But distortion works in both directions: not only do market participants operate with a bias, but their bias can also influence the course of events. This may create the impression that markets anticipate future developments accurately, but in fact it is not present expectations that correspond to future events but future events that are shaped by present expectations. The participants' perceptions are inherently flawed, and there is a two-way connection between flawed perceptions and the actual course of events, which results in a Jack of correspondence between the two. I call this two-way connection 'reflexivity' (Ibid., 14).

In essence, Soros argues that SMP are inherently volatile and unstable "since market participants are trying to discount a future that is itself shaped by market expectations" (Ibid., 314–315). As a consequence, the RMH conceptualizes and models SMP by incorporating two basic features (Ibid., 49):

(1) Markets are always biased in one way or another.
(2) Markets can influence the events that they anticipate.

(1) and (2) imply that equilibrium, rational investor, demand and supply, are not adequate or exhaustive concepts for interpreting the data and the behaviour of price trajectories—at least not in the ones offered by the EHM.

As regards the concept of equilibrium, the RMH states that there are "occasions when the bias affects not only market prices but also the so-called fundamentals. This is when reflexivity becomes important. It does not happen all the time but when it does, market prices follow a different pattern. They also play a different role: they do not merely reflect the so-called fundamentals; they themselves become one of the fundamentals which shape the evolution of prices. This recursive relationship renders the evolution of prices indeterminate and the so-called equilibrium price irrelevant" (Ibid., 7).

As concerns the notion of rationally, Soros argues that real investor are most of the times in a state of incomplete understanding and asymmetric information, so that their decisions cannot be rational in the neoclassical sense.

As concerns the notions of demand and supply, the RMH points out that their curves are not independent, for "both of them incorporate the participants' expectations about events that are shaped by their own expectations. Nowhere is the role of expectations more clearly visible than in financial markets. Buy and sell decisions are based on expectations about future prices, and future prices, in turn, are contingent on present buy and sell decisions. To speak of supply and demand as if they were determined by forces that are independent of the market participants' expectations is quite misleading" (Ibid., 29). Accordingly, it is quite possible that events in the marketplace may affect the shape of the demand and supply curves, a fact incompatible with neoclassical view. In effect, "the demand and supply curves are supposed to determine the market price. If they were themselves subject to market influences, prices would cease to be uniquely determined. Instead of equilibrium, we would be left with fluctuating prices. This would be a devastating state of affairs. All the conclusions of economic theory would lose their relevance to the real world" (Ibid., 29). In this sense, the RMH agrees with a criticism that goes back to Sraffa (see [21]). In essence, a simple logical examination (see e.g. [16]) shows on one hand that supply curve is a non sequitur: the neoclassical view, based on maximization of profit, does not imply the usual supply curve, but another kind of curve. On the other hand, it shows that demand curve has a different shape than the one predicted by the neoclassical view.

The RMH's verbal description and model of SMP rejects the core idea of the EMH, i.e. that the trends we can observe in financial markets are merely temporary aberrations that will be eliminated in the long run by the fundamental forces of supply and demand. It is simply untenable, since it ignores the possibility of reflexive dynamics: "there can be no assurance that 'fundamental' forces will correct 'speculative' excesses. It is just as possible that speculation will alter the supposedly fundamental conditions of supply and demand" [18, p. 31].[14]

Soros offers a penetrating verbal conceptualization and descriptions of the nature and role of reflexivity in many financial markets. Two notable examples are stock markets and currency market (Ibid., pp. 50–68). In the former, Soros selects some 'observable' features (i.e. variables) of stock markets and shows their reflexivity and the consequent price behavior. These features are:

(1) prevailing bias;
(2) underlying trends.

(1) is the result of a simplified (but not simplistic) interpretation of stock markets, that is the idea that "markets have many participants, whose views are bound to differ. […] Many of the individual biases cancel each other out, leaving what I call the 'prevailing bias'" (Ibid., 50). What it makes legitimate and plausible to accept this feature is the fact that the individual perceptions "can be related to a common denominator, namely, stock prices. In other historical processes, the participants' views are too diffuse to be aggregated and the concept of a prevailing bias becomes little more than a metaphor. In cases a different model may be needed, but in the stock market the participants' bias finds expression in purchases and sales" (Ibid.). The biases can be broken down into classes: (i) the positive and (ii) the negative. If we describe stock prices simply as rising and falling, then a positive prevailing bias pushes prices to raise; when it works in the opposite direction, it is negative. This implies that rising prices are reinforced by a positive bias and falling prices by a negative one. So, "other things being equal, a positive bias leads

[14] Soros offers a nice example of this particular process: "In the normal course of events, a speculative price rise provokes countervailing forces: supply is increased and demand reduced, and the temporary excess is corrected with the passage of time. But there are exceptions. In foreign exchange, for example, a sustained price movement can be self—validating, because of its impact on domestic price levels. The same is true in the stock market where the performance of a stock may affect the performance of the company in question in a number of ways. And in examining the recent history of international lending we shall find that excessive lending first increased the borrowing capacity of debtor countries, as measured by their debt ratios, and then, when the banks wanted to be repaid, the debtor countries' ability to do so evaporated. Generally speaking, we shall find that the expansion and contraction of credit can affect the debtors' ability and willingness to pay" [18, p. 30].

to rising stock prices and a negative one to falling prices. Thus the prevailing bias is an observable phenomenon" (Ibid.).

(2) is a trend that influences the movement of stock prices whether it is recognized by investors or not. The influence on stock prices will, of course, vary, depending on the market participants' views. In sum, we have that:

- a trend is 'self-reinforcing' when stock prices reinforce the underlying trend;
- a trend is 'self-correcting' when stock prices do not reinforce the underlying trend.
- a prevailing bias is 'self-reinforcing' when stock prices reinforce the prevailing bias
- a prevailing bias is 'self-correcting' when stock prices do not reinforce the prevailing bias

This means that when a trend is reinforced, it accelerates. On the other side, when a bias is reinforced, the divergence between expectations and the actual course of future stock prices gets wider. Conversely, when it is self-correcting, the divergence gets narrower.

Soros' qualitative, verbal reasoning shows us that "the trend in stock prices can then be envisioned as a composite of the "underlying trend" and the "prevailing bias"" (Ibid.). (1) and (2) generate trends in stock market prices, but here it comes reflexivity. In light of Soros' account, stock prices, underlying trend and prevailing bias, in turn, are affected by stock prices (see Fig. 3):

The interplay between stock prices and the other two factors has no constant: what is supposed to be the independent variable in one function is the dependent variable in the other. Without a constant, there is no tendency toward equilibrium. The sequence of events is best interpreted as a process of historical change in which none of the variables—stock prices, underlying trend, and prevailing bias—remains as it was before. In the typical sequence the three variables reinforce each other first in one direction and then in the other in a pattern that is known, in its simplest form, as boom and bust" (Ibid.)

It is also possible to express reflexivity in a formal way. For instance, with P = price, B = bias, U = underling trends, $\downarrow$ = decrease, $\uparrow$ = increase, we have:

$P = f(U), P = f(B)$

and,

$B = f(P, U = f(P)$.

Within this model it is possible to express qualitative relations, e.g. $(B\downarrow + U\downarrow) \rightarrow \downarrow P$, or $(B\uparrow + U\uparrow) \rightarrow \uparrow P$. The former is the qualitative formalization of a 'boom', the latter of a 'bust'.

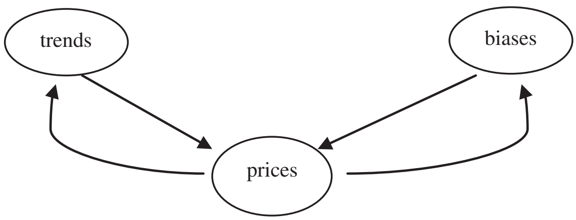


Fig. 3 Reflexivity at work: interplay between trends, biases and prices

In essence the RMH sets out to show the SMP are at great extent endogenous and "driven by positive feedback mechanisms involving investors' anticipations that lead to self-fulfilling prophecies" [3, p. 3]. In order to produce this finding, the RMH powerfully draws on a verbal conceptualization of financial phenomena. An example is the verbal report written by Soros [18, pp. 61–63] in February 1970 and titled "The case for mortgage trusts". In this report Soros offers a clear example of a verbal and qualitative reasoning about a financial process that cannot be handled by means of the standard (neoclassical) approach and that requires a new concept—reflexivity. In particular Soros argues that a self-reinforcing process can be generated by the interplay between future course of earning and price for those earnings. In the case of the mortgage trusts, he argued that three major factors reinforce each other in a reflexive way and on this basis he offered a scenario of the probable course of events, which eventually occurred. The reflexivity is the key factor in the process, as while "the conventional method of security analysis is to try and predict the future course of earnings and then to estimate the price that investors may be willing to pay for those earnings", Soros' contention is that in the case of "analysis of mortgage trusts because the price that investors are willing to pay for the shares is an important factor in determining the future course of earnings" (Ibid., 61).

Furthermore, reflexivity turns out to be a heuristic tool, as it promotes the formulation of new conjectures about possible relations between economic entities, which can be tested and can lead to new discoveries. In this sense Soros mention two interesting examples: the relation credit-collateral and regulators-economy.

In former example the RMH shows that neoclassical thesis—i.e. monetary values are a passive reflection of the state of affairs in the real world—is misleading: "money values do not simply mirror the state of affairs in the real world; valuation is a positive act that makes an impact on the course of events. Monetary and real phenomena are connected in a reflexive fashion; that is, they influence each other mutually" (Ibid., 18).[15] Soros' qualitative verbal model allows us to identify patterns in lending activity. In effect the reflexive interaction between the act of lending and collateral values shows that a "slowly accelerating credit expansion is followed by a short period of credit contraction—the classic sequence of boom and bust" (Ibid., 17). Moreover, the bust has a short course since "the attempt to liquidate loans causes a sudden implosion of collateral values" (Ibid., 18).

So for the RMH the explanation of bubbles and crashes is quite cogent and plausible: they are normal (and profitable, in Soros' case) outcomes of endogenous processes feed by reflexive relations between entities of stock markets—that is investors, their expectations and decisions, the prices. In particular, "in a boom/

[15] Soros uses the example of credit: "loans are based on the lender's estimation of the borrower's ability to service his debt. The valuation of the collateral is supposed to be independent of the act of lending; but in actual fact the act of lending can affect the value of the collateral. This is true of the individual case and of the economy as a whole. Credit expansion stimulates the economy and enhances collateral values; the repayment or contraction of credit has a depressing influence both on the economy and on the valuation of the collateral. The connection between credit and economic activity is anything but constant—for instance, credit for building a new factory has quite a different effect from credit for a leveraged buyout" [18, p. 18].

bust sequence we would expect to find at least one stretch where rising prices are reinforced by a positive bias and another where falling prices are reinforced by a negative bias. There must also be a point where the underlying trend and the prevailing bias combine to reverse the trend in stock prices" (Ibid., 51). This tipping point is the main problem for the RMH, for it does not provide a quantitative framework to detect it. It cannot tell when to take the crucial decision to invest/disinvest an amount of money. The lack for a quantitative answer to this particular problem, and the attempt to provide it, is the starting point of another hypothesis put forward to deal with the SMP problem, i.e. the econophysic hypothesis, or better the Far-form-Equilibrium-Hypothesis (FEH), based on the notion of far-from-equilibrium.

6 Far-from-Equilibrium Hypothesis

The FEH sets out to answer both the qualitative and quantitative side of the SMP problem. It draws on the findings of the hypotheses considered in so far and extends them with new features taken from physics and psychology. In effect the model provided by Soros plays an important role in the development FEH, which explicitly draws on the RMH and improves it under several respects, in particular the identification of the tipping point in a quantitative and predictive way. First of all Soros argues that reflexivity is the key factor in the production of far-form-equilibrium states in stock markets. In effect he maintains that it is possible to distinguish between:

(a) near-equilibrium conditions, "where certain corrective mechanisms prevent perceptions and reality from drifting too far apart" [18, p. 6];
(b) far-from-equilibrium conditions, "where a reflexive double-feedback mechanism is at work and there is no tendency for perceptions and reality to come close together without a significant change in the prevailing conditions, a change of regime" (Ibid.).

He points out that the Neoclassical theory applies only to (a), for "the divergence between perceptions and reality can be ignored as mere noise", while for (b) "the theory of equilibrium becomes irrelevant and we are confronted with a one-directional historical process where changes in both perceptions and reality are irreversible. It is important to distinguish between these two different states of affairs because what is normal in one is abnormal in the other" (Ibid.). Boom-and-bust is a typical example of a far-from-equilibrium process: Soros argues that it occurs because of the reflexivity between expectation of investor and prices.

So, while the EMH reduces the price fluctuations in SMP to rational reactions of the investors to the external and random arrival of new information affecting the future prospects of companies, Soros shows how price fluctuations can be generated by pure internal dynamics. In other words the EMH provide an exogenous explanation of price trajectories by means of the random arrival of external economic news, and the RMH and FEH offer an endogenous explanation (i.e. today's market price is a reaction to yesterday's prices change). In essence the FEH admits that 'news' about stock markets include the "most recent movements of stock prices themselves. In fact, in today's stock market, the major news will always be the most recent movements in stock prices, rather than 'real' news from the economy" [16, p. 342].

The FEH advances an answer to a crucial question about SMP posed by Soros [18, p. 9]: "how can near—and far-from-equilibrium conditions be distinguished from each other? What is the criterion of demarcation?". In order to do that, the FEH relies on a strong analogy between stock markets prices and critical points studied in statistical physics. In effect, a version of the FEH (see e.g. [7]) is based on the use of 'singularities' in the understanding of far-from-equilibrium behaviour. A critical point-like feature of SMP is 'herd' or 'crowd' behavior: "herd behavior is often said to occur when many people take the same action, because some mimic the actions of others" [7, p. 94]. Herding can be used to improve the FMH and RMH in the understating of SMP and to generate a quantitative model for it. In effect Sornette points out that the peculiar dynamics of confidence, contagion and decision making with imperfect information allow us to shed light on critical issues such as the mechanisms underlying crashes, the forecast of crashes, their control, the fundamental instability in the world financial structure.

The FEH shows that a cooperative behaviour among traders imitating each other can be expressed by means a log-periodic structure of the time evolution of the system. This fact implies that, in a precise sense, the market anticipates some critical events, like a crash, in a subtle self-organized and cooperative fashion, by "releasing precursory 'fingerprints' observable in the stock market prices. In other words, this implies that market prices contain information on impending crashes. If the traders were to learn how to decipher and use this information, they would act on it and on the knowledge that others act on it; nevertheless, the crashes would still probably happen" (Ibid., 280).

In this sense the FEH supports a weaker form of the EFM, "according to which the market prices contain, in addition to the information generally available to all, subtle information formed by the global market that most or all individual traders have not yet learned to decipher and use" [7, p. 280]. The FEM challenges the reading of the EMH according to which traders extract and consciously incorporate in their action all information contained in the market prices, and proposes that the market as a whole exhibit 'emergent' behavior.

An emergent behavior like herding affects the systems in a deep and lasting way, creating bifurcations and phases that can be forecast. In effect, the FEH draws on the idea that "we are much more interested in forecasting the major bifurcations ahead of us, involving the few important things, like health, love, and work, that count for our happiness" (Ibid.). Similarly, the prediction of the detailed evolution of a complex systems

has no real value, and the fact that we are taught that it is out of reach from a fundamental point of view does not exclude the more interesting possibility of predicting phases of evolutions of complex systems that really count, like the extreme events. It turns out that most complex systems in natural and social sciences do exhibit rare and sudden transitions that occur over time intervals that are short compared to the characteristic time scales of their posterior evolution. Such extreme events express more than anything else the underlying 'forces' usually hidden by almost perfect balance and thus provide the potential for a better scientific understanding of complex systems" (Ibid.).

Thus the FEH relies on the finding that the long-term behavior of these complex systems is often shaped by rare catastrophic events (e.g. the large earthquakes that happen in California every two centuries or so account for a significant part of the total tectonic deformation). In the same way the psychological state of investors will be shaped by a financial crash that might destroy trillions of dollars instantaneously.

More specifically, Sornette points out that "the organization of spatial and temporal correlations do not stem, in general, from a nucleation phase diffusing across the system. It results rather from a progressive and more global cooperative process occurring over the whole system by repetitive interactions" (Ibid., 19)—like herding. Therefore, the FEH applies these ideas to the SMP, in particular financial crashes, and combines findings and tools from mathematics, physics, engineering, and the social sciences in order "to identify and classify possible universal structures that occur at different scales and to develop application-specific methodologies for using these structures for the prediction of the financial 'crises.' Of special interest will be the study of the premonitory processes before financial crashes or 'bubble' corrections in the stock market" (Ibid., 20). Thus, the FEH offers a new interpretation of crashes and bubbles: while classical models propose mechanisms at very short time scales in order to account for a collapse, the FEH proposes that they have to be searched at long time scales, and that the progressive increase of the market prices is produced by an increasing build-up of the market cooperation. In this light the issue of the particular way by which prices collapse is not central, since "according to the concept of the critical point, any small disturbance or process may have triggered the instability, once ripe. The intrinsic divergence of the sensitivity and the growing instability of the market close to a critical point might explain why attempts to unravel the local origin of the crash have been so diverse. Essentially all would work once the system was ripe" (Ibid., 280). The bottom line: a crash has an endogenous origin and "exogenous shocks only serve as triggering factors" (Ibid.). The origin of the crash is much more subtle, deep and it is generated progressively by the whole market. The FEH argues, on the basis of the analogies, that stock market crashes are caused by the slow build-up of long-range correlations leading to a global cooperative behavior of the market and eventually ending in a collapse in a short, critical time interval, that around a critical point defined as the explosion to infinity of a normally well-behaved quantity. A crash may be triggered by local self-reinforcing imitation between traders. If this tendency for traders to 'imitate' their 'friends' increases up to a certain point called the 'critical' point, many traders may place the same order (sell) at the same time, thus causing a crash.

Tellingly the FEH advances an answer also to the quantitative side of the SMP problem, with the thesis the herding behavior and its critical point can be described by a log-periodic formula that approximates the data and offers a basis for prediction of an incoming crash.[16] In addition, the endorsement of the FEM does not imply a complete dismissal of the EMH, for "our results suggest a weaker form of the 'weak efficient market hypothesis', according to which the market prices contain, in addition to the information generally available to all, subtle information formed by the global market that most or all individual traders have not yet learned to decipher and use" (Ibid., 279). This result relies an emergentist view, which modifies but do not reject the EHM: in effect the emergentism provides a new statistical reading of the EMH as "instead of the usual interpretation of the efficient market hypothesis in which traders extract and consciously incorporate (by their action) all information contained in the market prices, we propose that the market as a whole can exhibit "emergent" behavior not shared by any of its constituents" (Ibid.).

Of course the FEH does not tell that a crash is a deterministic outcome of a bubble, so that "it remains rational for investors to remain in the market provided they are compensated by a higher rate of growth of the bubble for taking the risk of a crash, because there is a finite probability of 'landing smoothly,' that is, of attaining the end of the bubble without a crash" (Ibid., 25). It maintains that even if a crash is not a certain result of a bubble, it can be signaled by precursors.

Moreover this model can be put to use to search for endogenous precursor of crashes that occurs also at very small interval of time, like a flash-crash or a mini flash-crash. It is worth noting that especially in stock markets the relations between a prediction and the system to which it is directed is highly problematic. In effect, as emphasized by Sornette himself, after a hypothesis has been formulated (say a prediction a crash of an amplitude between 15 and 22 % will occur between one and three months from now) there at least three different outcomes:

(a) nobody believes the prediction.
(b) Everybody believes it.
(c) Sufficiently many investors believe that it may be correct.

In case (a) the prediction, which is useless since it does not generate a reaction, even if it is correct and the market crashes, is not considered a victory for the theory since some criticism can point out that is just a "lucky one", with no statistical relevance. In case (b) the prediction generates panic, and accordingly the market crashes. Thus, the prediction seems self-fulfilling and the criticism can point out that the success is due to the panic effect rather than to its predictive power. In case (c) the investors make reasonable adjustments, and accordingly the bubble vanishes and there is no crash: the prediction hence disproves itself. It follows that all these outcomes are embarrassing for the theory: "in the first two, the crash is not avoided, and in the last scenario the prediction disproves itself and as a consequence the theory looks unreliable. This seems to be the inescapable lot of scientific investigations of systems with learning and reflective abilities, in contrast with the usual inanimate and unchanging physical laws of nature" (Ibid, 33).

---

[16] In particular Sornette provides the following mathematical model: $F_{lp}(t) = A_2 + B_2\,(t_c - t)^{m2}$ $[1 + C\cos(\omega\log((t_c - t)/T))]$ [7, p. 232].

6.1 Forecasts of Flash-Crashes

Recent findings about the SMP show that events like a crash are not rare. Moreover, they do not happen only at large time scales, but continually occur also at small or very small time scales (even seconds). SMP continually experience bubbles, crashes and flash-crashes: it is plenty of these far-from-equilibrium events. The point is that the FEH can be usefully applied in the detection of signs and marks of incoming flash crashes or mini flash crashes.[17] Flash crashes are abrupt and severe price changes (falls) that occur in an extremely short period. The most famous flash crash is the May 6th 2010 Flash Crash (see Fig. 4): the index sank 900 points in less than 5 m, but recovered almost all the losses in the next 15 m of trading.

A popular example of mini flash crash , which occurs at even smaller time scales, is the one occurred to Apple stock on Jan. 25, 2013 (Fig. 5).[18]

Drawing on the verbal model provided by Soros (RMH) based on endogenous dynamics of the SMP and quantitative models of non-linear systems, Sornette has developed a formal tool that aims at forecasting bubbles and crushes by means of the detection of level of endogeneity in SMP. Such a level is "used for characterizing the robustness of systems and for developing diagnostics of fragility and of incoming crises as well as upside potentials" [3, p. 2]. Since financial markets can be seen as nothing but the engines through which information is transformed into prices, if the information and news incorporated by prices are the one concerning the recent movements of prices, then that may lead to deviations from true valuation, generating a crash or a bubble. Drawing on this idea Sornette has created a measure of the level of endogeneity "which reflect a robust behavioral trait of human beings who tend to herd more at short time scales in time of fear and panic. Our study thus complements the evidence for herding at the time scales of years over with financial bubbles develop, by showing the existence of herding at short time scales according to a different mechanism than the ones operating at large time scales" (Ibid., 4). These mechanism can be captured and detected by the FEH, which is then incorporate domain-specific features of SMP into a new qualitative and formal model.

[17] Mini Flash Crashes, or Flash Equity Failures, were first identified by Nanex LLC. To qualify as a down crash, the stock price change has to satisfy the following conditions (see [23, 24]):
(i) it has to tick down at least 10 times before ticking up,
(ii) price changes have to occur within 1.5 s,
(iii) price change has to exceed -0.8 %.
The same holds for an up crash candidate, but in the opposite direction.

[18] The stock fall about 2 % in the last minute of trading, with 1 million shares exchanged. It amounts to 10 times the volume during any other time that day, and the drop evaporated nearly $7 billion of Apple's market value. In the final few seconds of trading, Apple recovered more than half of that.

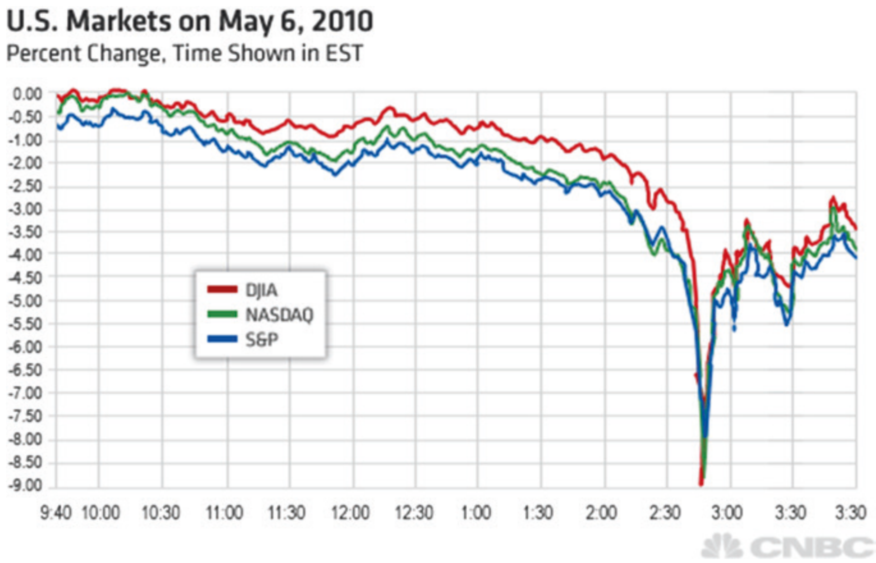


Fig. 4 The 6th May 2010 flash crash

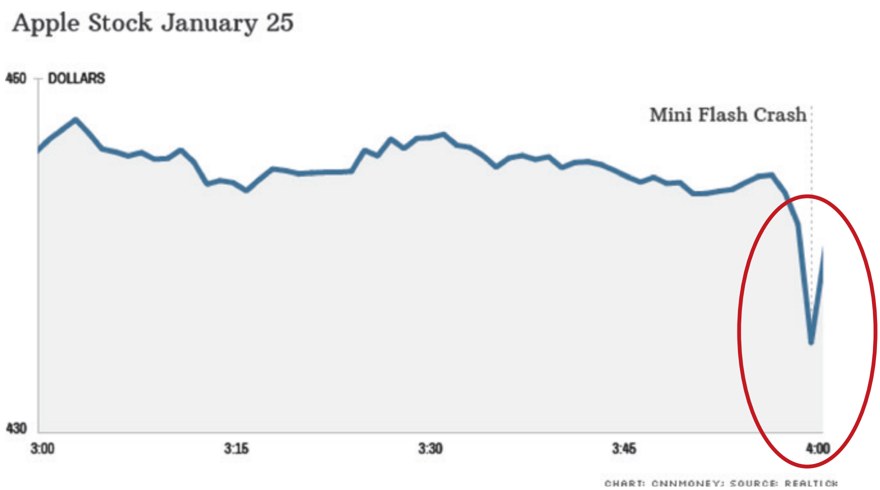


Fig. 5 Apple stock mini-flash crash

7 Ways of Generating Hypotheses: A Comparison

Notwithstanding the difficult inter-relation between the predictions about the SMP drawn from a certain hypothesis and the SMP's behavior itself, we can note that, by criticizing the EMH, the other three hypotheses develop a line that offers better and better answers both to the qualitative and the quantitative side of the SMP problem, in particular the explanation and forecast of bubbles and crashes. Additionally, a comparison of the ways for generating these four hypotheses reveals us something important about their efficiency and plausibility. The examination of these ways shows why some of these hypotheses are successful and efficient and some not, and also shed light on the extent to which a particular hypothesis can be usefully applied. Thus, the study of the means of generation of hypotheses offer us a guide for formulating new hypotheses in a plausible and rational fashion. More specifically I argue that the examples discussed above tells us that the an effective generation of a hypothesis has to draw on a preliminary verbal conceptualization (a discourse) on the data of a specific subject, that is a verbal description of it, which establishes the entities to investigate, their properties and relations, and a set of variables that affect them. This is a bottom-up process and it is the only way to incorporate the domain-specific features of the subject in a plausible representation of it, which can end up with a mathematical theory. Thus my thesis[19] is that generation of genuinely new hypotheses and new mathematics (formal treatments) stem from such a preliminary verbal modeling and conceptualization, which delimitate the features of the phenomenon at stake and their relations. Such a preliminary conceptualization, as verbal, often offers a hybrid and multivalent representation (see [29]) of the subject of inquiry: concepts and tools from other domains are incorporated in the conceptualization by means of ampliative reasoning—like analogies and metaphors. In effect, similarities and dissimilarities between this and other phenomena have to be recognized and integrated at this stage of the conceptualization (e.g. the analogy between earthquake and crash in the FEH). After a verbal conceptualization is generated, in order to formulate a hypothesis a selection of features of the phenomenon (in our case the SMP) is made for incorporation in a theory. This selection may be expressed mathematically in most of the cases.

More precisely, the generation of hypotheses can be broken down into four parts:

(1) starting from the data (bottom-up), a verbal and informal reasoning generates an account for the phenomenon at stake—i.e. a verbal description that includes and highlights certain features of the phenomenon, and ignores or deliberately cuts off others.[20]
(2) Such a description leads to an informal conceptualization, where:

- relations and entities, and their possible hierarchies, are clarified;
- the reasons pro and contra the choice for taking into account some variables and not others are specified;
- similarities and dissimilarities with other known phenomena are put forward.

[19] The study of ways for generating hypotheses and ampliating knowledge has reignited in the last 40 years (see e.g. [2, 17, 25–36]). My work is indebted to these approaches, but it tries to push forward the frontier of research.

[20] Obviously such a description is, in turn, theory-laden, since it must rely on previous conceptualization and, if available, formal theories about the constituents of the phenomenon—but I will not treat this problem here.

(3) This informal conceptualization produce a theory, which gives the principles and the local or general laws underpinning the phenomenon. These principles and laws are hypotheses, even though plausible and grounded on findings.
(4) In the best scenario, a new formal quantitative treatment is generated (a new mathematics—e.g. calculus, fractal mathematics, path integrals). The theory can be expressed in a mathematical form and the generation of hypotheses ends up with a formal-quantitative treatment.

We can now see whether, how and to what extent the four different hypotheses considered in the paper fit this model (M), and above all whether and how their effectiveness is related to it. In effect, the hypotheses about the SMP examined in the paper show different ways of generation.

First of all the EMH does not fit this model at all: it does not stem from the attempt to provide a verbal conceptualization of the SMP behavior, capable to reflect the peculiar features of SMP. Instead, it aims at introducing an idealized account for it that is treatable by known mathematics and is generated from a pre-existing description offered for physics—it is based on the 'mechanical analogy' (see [37]). This description and its formal tools are transferred to a suitably selected part of the data about SMP. I argue that this explains why the EMH is not so effective in dealing with the SPM. In effect, the EHM is generated in open violation of the model M: there is no attempt to build a theory by means of a cogent and plausible conceptualization of all data and the peculiar features of the SMP. On the contrary, no specific features are incorporated in the hypotheses and there is no descriptive analysis of the SMP—it can be viewed as a normative account. Sure, the EMH is not completely flaw: it relies on ampliative inferences (the mechanical analogy), but mechanical properties are superimposed to data—not generated from the data. Thus the EMH can be effective only in those (few) cases in which the analogy between mechanics and finance works. But it means to say that the SMP features are the same as the ones underlying the mechanical processes, which is at least highly controversial. In addition, the EMH is generated in a static way, that is top-down, axiomatically, trying to save a given framework and to extend a pre-existing formal tools only to a part of the data (see again [17] on this point). So it is not surprising that the main versions of the EMH are flawless and unreliable. Thus, the way by which the EMH is generated, which I have labelled 'static', explains most of its weakness.

On the other hand the FMH, RMH and FEH are generated in a different way and also show instructive differences between each other's. These differences explain their different effectiveness.

The FHM draws on the findings about few peculiar features of the SMP and reflects these features in a new mathematical model, telling a quite different story than EMH about the SMP. More precisely, Mandelbrot produced a hypothesis that takes into account features like discontinuity, dependence, concentration and scaling that are largely ignored or explicitly denied by the EFM. The FMH is an attempt to provide a new quantitative answer to SMP by arguing that the neoclassical equilibrium and efficiency are not characteristics of the SMP. The FMH employs a bottom-up approach: it starts from a collection of data about the SMP and to make sense of it by introducing concepts and properties that are able to describe the inner relations between data. Even if these properties are not good for forecasting, they open up new readings of the SMP. In particular they allow to make sense of instability and to pose the problem of the role of endogenous forces in SMP.

The RMH incorporates in a verbal, qualitative model, a critical feature of SMP, i.e. investors' expectations and the consequent reflexive and endogenous behavior. In a sense, it fills some of qualitative gaps in the RMH, even if it does not offer a quantitative model of the SMP behavior, in particular about the forecast of the tipping points. It powerfully employs a bottom-up approach: the description of some critical financial processes (e.g. Soros' report) stems from an analysis of the critical events in the SMP, like bubbles and crashes, and builds the concepts needed to make sense of it. These concepts, i.e. reflexivity and endogeneity, provide a new reading of the SMP behavior, which accounts for critical events that shape the behavior of the SMP for long periods.

Finally, the FEH refines some of the main features of the FMH and RMH by modeling the process of herding and the use of the bifurcation theory. In this way it offers a quantitative and qualitative reading of the SMP features that allow us to produce interesting statistical forecasts. In effect, it offers a new statistical interpretation of the SMP and in order to do that it mixes up different concepts from distinct fields (psychology and physics) and a careful interpretation of a big amount of data about SMP in critical points (i.e. crashes and bubbles). Sure, it can't provide a tools to make predictions like in physics, but it sets out to foresee the fingerprints of these critical points.

Tellingly, the last three hypotheses build upon each other, or better upon the same web of concepts. They recognize and model crucial features of the SMP that are neglected or deliberately cut off in the EMH—such as concentration, scaling, endogeneity, herding. Similarly, their ways of generation differ a bit, but in essence are on the same line and contribute to produce better and better readings, analysis and, when possible, forecast about the SMP. This line is shaped by a 'dynamic' approach: bottom-up, local, and that sets out to reflect and incorporate domain-specific features.

References

1. Cellucci, C.: Filosofia e Matematica. Laterza, Roma-Bari (2002)
2. Cellucci, C.: Rethinking Logic. Logic in Relation to Mathematics, Evolution, and Method. Springer, New York (2013)
3. Sornette, D., Filimonov, V.: Quantifying reflexivity in financial markets: towards a prediction of flash crashes. Phys. Rev. E **85**, 056108 (2012)
4. Sornette, D., Kaizoij, T.: Market Bubbles and Crashes. arXiv:0812.2449 [q-fin.RM] (2008)
5. Mandelbrot, B.: The Misbehavior of Markets: A Fractal View of Financial Turbulence. Basic Book, New York (2006)
6. Silver, N.: The Signal and Noise. The Penguin Press, New York (2012)
7. Sornette, D.: Why Stock Markets Crash. Princeton University Press, Princeton (2003)
8. Haugen, R.A.: The Inefficient Stock Market. Prentice-Hall, New Jersey (1999)

9. Minsky, H. (1982 [1963]) Can 'It' Happen Again? Essays on instability and finance, Armonk. NY: M. E. Sharpe
10. Galbraith, J.K.: The Great Crash 1929. Houghton Mifflin Company, Boston (1954)
11. Kindleberger, C.P.: Manias, Panics and Crashes: A History of Financial Crises. Basic Books, New York (1978)
12. Ferguson, : The Ascent of Money. Penguin Press, New York (2008)
13. Friedman, M.: The methodology of positive economics. In: Caldwell, B. (ed.) (1984) Appraisal and Criticism in Economics: A Book of Readings (Reprinted). Allen & Unwin, London (1953)
14. Fama, E.F.: Efficient capital markets: a review of theory and empirical work. J. Finance **25**(2), 383–417 (1970)
15. Fama, E.F.: Efficient capital markets II. J. Finance **46**(5), 1575–1617 (1991)
16. Keen, S.: Debunking Economics. Zed Books, New York (2011)
17. Ippoliti, E.: Generation of hypotheses by ampliation of data. In: Magnani, L. (ed.) Model-Based Reasoning in Science and Technology. Theoretical and Cognitive Issues. Springer, Heidelberg/Berlin (2013)
18. Soros, G.: The Alchemy of Finance. Wiley, New York (2003)
19. Mirowski, P.: More Heat than Light: Economics as Social Physics: Physics as Nature's Economics. Cambridge University Press, Cambridge (1989)
20. Mandelbrot, B.: The Variation of Certain Speculative Prices. J. Bus. **36**, 394 (1963)
21. Peters, E.E.: Fractal Market Analysis. Wiley, New York (1994)
22. Sraffa, P.: The Production of Commodities by Means of Commodities: Prelude to a Critique of Economic Theory. Cambridge University Press, Cambridge (1960)
23. Nanex: Nanex flash crash summary report. http://www.nanex.net/FlashCrashFinal/FlashCrashSummary.html (2010)
24. Nanex: Flash crash analysis—final conclusion. http://www.nanex.net/FlashCrashFinal/FlashCrashAnalysisTheory.html (2010). 6 May 2010
25. Lakatos, I.: Proofs and Refutations. Cambridge University Press, Cambridge (1976)
26. Nickles, T. (ed.): Scientific Discovery. Logic and Rationality. Reidel, Netherlands (1980)
27. Nickles, T. (ed.): Scientific Discovery: Case Studies. Reidel, Netherlands (1980)
28. Nickles, T., Meheus, J. (eds.): Models of Discovery and Creativity. Springer, New York (2009)
29. Cellucci, C., Gillies, D. (eds.): Mathematical Reasoning and Heuristics. College Publications, London (2005)
30. Gillies, D.: Revolutions in Mathematics, pp. 353 + xi. Oxford University Press, Oxford (1992) (Editor and Contributor)
31. Kantorovich, A.: Scientific Discovery: Logic and Tinkering. Suny Press, New York (1993)
32. Groshoz, E.: Representation and Productive Ambiguity in Mathematics and the Sciences. Oxford University Press, New York (2007)
33. Grosholz, E., Breger, H.: The Growth of Mathematical Knowledge. Springer, New York (2000)
34. Magnani, L.: Abduction, Reason, and Science. Processes of Discovery and Explanation. Kluwer Academic/Plenum Publishers, New York (2001)
35. Magnani, L., Nersessian, N.J. (eds.): Model-Based Reasoning. Scientific Discovery, Technological Innovation, Values. Kluwer Academic/Plenum Publishers, New York (2002)
36. Magnani, L., Nersessian, N.J., Thagard, P. (eds.): Model-Based Reasoning in Scientific Discovery. Kluwer Academic/Plenum Publishers, New York (1999)
37. McLure, M.: Pareto, Economics and Society: The Mechanical Analogy. Routledge, London, New York (2001)