



Scientific innovation: A conceptual explication and a dilemma

(La innovación científica: una explicación conceptual y un dilema)

Thomas STURM*^{1,2}

¹ ICREA, Spain

² Universitat Autònoma de Barcelona, Spain

ABSTRACT: I offer an analysis of the concept of scientific innovation. When research is innovated, highly novel and useful elements of investigation begin to spread through a scientific community, resulting from a process which is neither due to blind chance nor to necessity, but to a minimal use of rationality. This, however, leads to tension between two claims: (1) scientific innovation can be explained rationally; (2) no existing account of rationality explains scientific innovation. There are good reasons to maintain (1) and (2), but it is difficult for both claims to be accepted simultaneously by a rational subject. In particular, I argue that neither standard nor bounded theories of rationality can deliver a satisfactory explanation of scientific innovations.

KEYWORDS: Scientific innovation, rationality, heuristics, models of scientific change, science policy.

RESUMEN: Ofrezco aquí un análisis del concepto de innovación científica. Cuando se innova en la investigación, los elementos de investigación altamente novedosos y útiles comienzan a extenderse a través de una comunidad científica como resultado de un proceso que no se debe ni al ciego azar ni a la necesidad, sino a un uso mínimo de la racionalidad. Esto, sin embargo, genera tensiones entre dos afirmaciones: (1) la innovación científica puede ser explicada racionalmente; (2) ninguna explicación existente de la racionalidad da cuenta de la innovación científica. Hay buenas razones para mantener (1) y (2), pero es difícil que ambas afirmaciones sean aceptadas simultáneamente por un sujeto racional. En particular, sostengo que ni la teoría estándar de la racionalidad ni la teoría de la racionalidad acotada pueden ofrecer una explicación satisfactoria de las innovaciones científicas.

PALABRAS CLAVE: Innovación científica, racionalidad, heurística, modelos de cambio científico, política científica.

* **Correspondence to:** Thomas Sturm. ICREA, Pg. Lluís Companys 23 (08010 Barcelona), Spain. Dept. de Filosofia, Universitat Autònoma de Barcelona (08193 Bellaterra-Barcelona), Spain. – Thomas.Sturm@uab.cat – <http://orcid.org/0000-0001-5698-3542>

How to cite: Sturm, Thomas. (2019). «Scientific innovation: A conceptual explication and a dilemma»; *Theoria. An International Journal for Theory, History and Foundations of Science*, 34(3), 321-341. (<https://doi.org/10.1387/theoria.20652>).

Received: 13 March, 2019; Final version: 11 September, 2019.

ISSN 0495-4548 - eISSN 2171-679X / © 2019 UPV/EHU



This article is distributed under the terms of the
Creative Commons Attribution 4.0 International License

I. *Setting the agenda: a concept widely used though much neglected*

Scientists are often asked to promote innovation and aid society by, for instance, novel drugs and therapies, means of communication, ways of making technical devices more energy efficient, or methods for teaching mathematics to schoolchildren. Increasingly, they are also invited (if not urged) to innovate science itself. Academic institutions, grant agencies, and governments all encourage researchers to devise novel questions, methods, concepts, models, theories, goals, instruments, and even research institutions. But while the terminology of innovation is now widely used, all too often this is purely rhetorical rather than reflective. Here, I aim to foster philosophical debate concerning scientific innovation. The scientific system should be self-critical when it comes to the language required for, and used in, grant applications, research practice guidelines and reports of results, as well as the public dissemination and assessment of research. It should be clear, with all the necessary caveats, what we mean when we qualify research as “innovative”.

For starters, consider the following examples. The National Endowment for the Humanities (NEH), an independent US Government agency, offers Digital Humanities Advancement Grants, “leading to innovative work that can scale to enhance research, teaching, and public programming in the humanities.” It also aims to encourage “innovative collaborations between museum or library professionals and humanities professionals to advance preservation of, access to, use of, and engagement with digital collections and services.”¹ Again, the German Max Planck Society (MPG) takes up “new and innovative areas of research to supplement research carried out by German universities.”² Whenever a gap in the academic system is detected—a new interdisciplinary research field or discipline not yet represented at German universities, or an approach within a discipline not sufficiently explored—the MPG considers creating an institute in order to train a new generation of researchers who will be ready to fill emergent university chairs. Thus, the MPI for the Science of Human History (Jena, founded in 2014) uses evolutionary theory and biomolecular methods to research historical phenomena such as the Black Death in early modern Europe; while it also integrates biology and linguistics in the study of the history of human languages. Such natural-scientific approaches to human history are certainly unconventional.³

Finally, the European Union’s 8th “Framework Programme for Research and Innovation”, also called “Horizon 2020”, will have been funded to the tune of some €77 billion over 7 years (2014-2020; its successor, *Horizon Europe*, 2021-2027 has a budget of around €100 billion)⁴. Its stated aim is to break “down barriers to create a genuine single market for knowledge, research and innovation.”⁵ The quest for innovation for industries and to

¹ See <https://www.neh.gov/grants/odh/digital-humanities-advancement-grants>. – For the US National Science Foundation, see Nickles, 2015, pp. 69-71.

² See <http://www.germaninnovation.org/about-us/gcri-partner-institutions/max-planck-gesellschaft>

³ Currently (2019) more than 80 MPIs exist. While more than 40 Institutes have been closed since the MPG started its work, shutting one down is difficult because they develop a life of their own. So, closing sometimes also takes the form of integrating institutes or rebranding them.

⁴ See https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme_en

⁵ See <http://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>

meet societal challenges (such as healthcare, climate change, or food security) plays a central role here, of course. But *Horizon 2020* also has a €24.4 billion Excellent Science section,⁶ devoted to “the most promising avenues at the frontier of science” (via ERC grants), and to “excellent and innovative research training” (the so-called Marie Skłodowska-Curie Actions), among other things.⁷ Here, proposals can be submitted in any area, with no demand for marketable products or immediate social applications. Researchers are instructed to submit proposals that satisfy three basic requirements: (i) they should innovate research itself by means of so-called “high-risk” and “high-gain” ideas; but they must also demonstrate (ii) the importance and (iii) the feasibility of these innovative ideas. You just cannot seem to come up with ideas that are too outlandish when it comes to pursuing a research topic! Preliminary evidence that your research project is relevant and feasible must accompany proposals; but high-risk, high-gain ideas differ across disciplines. In the humanities, you are recommended to present a major novel concept to study your subject. In the natural sciences, the frequent advice is to introduce an innovative method or instrument to tackle unsolved problems. If you do not deliver along these lines, the ERC may not only reject your proposal but also exclude you from submitting a proposal for one or more subsequent years. In part, this depends on how well you balance the requirements mentioned.

So we can see that the language of innovation has moved from technology and the economy into science itself. However, none of the agencies or institutions tells us what innovative science really is; or, more importantly, what rules and tools can and should be used to foster and assess innovation.⁸ Evaluation panels and experts are in many ways left

⁶ http://ec.europa.eu/research/horizon2020/pdf/press/fact_sheet_on_horizon2020_budget.pdf

⁷ See <http://ec.europa.eu/programmes/horizon2020/en/h2020-section/excellent-science>

⁸ One attempt to solve this problem, at least partially, is to mechanize assessments or decision-making by introducing *quantitative* criteria, e.g.: measuring the applicant’s previous research output, especially in terms of normatively laden classifications of scientific journals (e.g., Q1-Q4 of the Web of Knowledge), numbers of citation, or funding previously obtained. This, however, measures *past*, not future performance. Past performance should not be ignored, but the attempt to mechanize it by quantitative measures has various disadvantages. For instance, the system incentivizes scientists to publish as many papers in higher journal categories in as short a time as possible, which can be more easily achieved by non-innovative contributions, thus potentially pushing aside innovative work that requires higher investments into time and cognitive resources. Similarly, standards of journal quality or the significance of citation numbers can differ from field to field: thus, in large disciplines with many journals, some journals impose extremely demanding standards, whereas in some very small disciplines even the leading journals are more flexible. Thus, Q1 in one discipline need not indicate the same standard as in another. Also, in psychology, a citation very often indicates approval, whereas in e.g., analytic philosophy citations often indicate clear-cut rejection or criticism, or at least more often. Another risk lies in large grants for young researchers (such as ERC starting grants): they are necessarily given to researchers at early stages of their careers, whose CVs do not yet provide, so to speak, statistical data of adequate sample size. Absent good sample sizes, the possibility of error and bias increases. Since each year only a few ERC starting grants are handed out, other young researchers with CVs that are (nearly) as strong as those of the winners are left behind. Also, winning such a grant reinforces winners’ careers, which again disproportionately favors them over those who had equally good, or perhaps better proposals, but did not win (this follows the well-known *Matthew effect*: “For to every one who has will more be given, and he will have abundance; but from him who has not, even what he has will be taken away”; *Matthew* 25:29; cf. Merton, 1968). Such considerations speak in favor of a more equal distribution of

to their own devices—their expertise, traditions, methods and values. These are often domain-specific: what counts as innovation in one area (say, the life sciences) is determined by criteria that are often irrelevant or ignored in other areas (e.g., the social sciences or humanities). Researchers who are being evaluated are confronted with a lack of clarity and cogency concerning standards of innovativeness.

Several important questions naturally arise here. The first is the obvious one: What *is* scientific innovation? Let us call this the *conceptual* question. Furthermore (the *explanatory* question): How should we explain the emergence of scientific innovation? Then (the *evaluative* question): How should we evaluate claims of innovativeness? And finally (the *science policy* question): How can we foster innovative research projects? Although there are many studies of innovation in general, very few of them deal with the specific topic of *scientific* innovation, let alone these key questions.

This might have been different. Before *The Structure of Scientific Revolutions* first appeared in 1962, in 1959 Thomas Kuhn published an article titled: “The Essential Tension: Tradition and Innovation in Scientific Research” (Kuhn 1977). That was probably the first contribution to make an analytical use of the terminology of innovation in science studies.⁹ Kuhn claimed that science has to strike the right balance between two opposing modes of thinking: so-called “convergent” and “divergent” thinking. One must start from an established consensus and look for new directions. The convergent mode of thinking tries to fit recalcitrant data into a preferred theoretical framework; while the divergent mode of thinking aims to develop alternatives to the dominant framework. Kuhn claimed that both modes of thinking are equally important for science, from the perspective of its long-term success. This is the so-called “risk-spreading argument” (D’Agostino 2010, 6). One should divide up the roles between different scientists, thereby reaping the benefits of each mode of thinking while reducing risks. Kuhn’s point still appears valid. When individual researchers are invited to present innovative research proposals, they must combine both modes of thinking. As the *Horizon 2020* requirements state it, one can only innovate in science if one can show the importance of one’s idea relative to the present state of the art and, furthermore, the feasibility of one’s proposal. So, it is necessary to grasp accepted knowledge and the gaps in it, together with current methods and their potential as well as their limits, and thereby the plausible paths leading to answers to unsolved problems. Moreover, dedicating large amounts of financial support to innovative projects does not necessarily undermine the risk-spreading argument: large ERC grants are given to only a few, while the majority of researchers continue to do less costly work along more familiar lines. Of course, I am saying nothing here as to whether grant agencies always fund the right projects, or what the right ways of dividing budgets up among convergent and divergent modes of research are.

funding resources, especially at earlier career stages; another alternative might be the random distribution of funds (Gillies, 2008). (Note: I do not wish to argue that quantitative criteria are entirely useless, but they ought not to be used mindlessly or mechanically. At the very least, evaluation panels need to be informed about the risks and disadvantages of such criteria.)

⁹ A hint for further research: a search on Google Ngram reveals that the general English word ‘innovation’ was used from the Renaissance up until it was forgotten in the 18th century, but it became more popular again beginning in the 1950s.

Kuhn did *not* make the concept of scientific innovation popular. On the contrary, he did not develop it further in later writings.¹⁰ Furthermore, he did not tell us much about the conceptual, explanatory, normative, or historiographical questions that I mentioned and that the concept of research innovation so clearly invites. Accordingly, unlike ‘discovery’, ‘crisis’, ‘paradigm’, or ‘revolution’, and other concepts ‘innovation’ is often assimilated into or confused with,¹¹ the concept of innovation is not a developed analytical tool.¹²

In the remainder of this essay, I try to build the foundations for further work. First, I spell out minimal features of scientific innovation, thus addressing the conceptual question. Second, that analysis leads to a puzzle related to the three other questions. I will explain that puzzle, especially with respect to the explanatory question, and discuss some attempts at solving it. In my view, the puzzle partly explains why criteria of innovativeness are so hard to spell out. That should inspire us to be more cautious when using the term.

II. Four conceptual claims concerning scientific innovation

For reasons that are well known, it is difficult—if not impossible—to give a complete conceptual analysis of ‘scientific innovation’. It is doubtful that the phenomenon has any essence that we could identify in terms of necessary and sufficient conditions, certainly not from the armchair using thought experiments that would provide us with rational insight into that essence. Does the innovation that results from a new laboratory instrument used in cancer research have much in common with the blending of theories from different disciplines (which, for example, made the discovery of DNA possible), or with a novel method for the long-term preservation of the original copy of Thomas Jefferson’s *The Life and Morals of Jesus of Nazareth*? It would seem not. In addition, given that the concept of scientific innovation has grown out of practices of evaluating research, and that standards for this are notoriously hard to fix, we can expect people to mean quite different things when they use similar terminology. The terminology is far from exact, as will soon be made clear.

Alternatively, if ‘scientific innovation’ is ill suited for complete conceptual analysis, and if it is inexact, then one might think that an account of it needs to be given in terms of what Carnap (1950, 1-18) has called an “explication”. As he famously argued, inexact concepts such as ‘probability’ must be explicated in ways that reveal and rectify deficiencies of our ordinary understanding of them: one needs to work towards “the transformation of

¹⁰ The distinction between “convergent” and “divergent” thinking is not the same as the famous one between “normal” and “revolutionary” science that forms the backbone of *Structure*. Convergent thinking is important not only in normal science, but in revolutionary science as well. As Kuhn (1977, 227) wrote, “only investigations firmly rooted in the contemporary scientific tradition are likely to break that tradition and give rise to a new one”. Moreover, the distinction between “normal” and “revolutionary” science only applies to “mature” sciences, while that between convergent and divergent thinking applies to all sciences, and even beyond that to the arts, technology, and other domains that promote innovation.

¹¹ One such confusion—between innovation and revolution—can be found in Nickles (2013).

¹² Among the few exceptions are: Nickles 2003, 2015, 2016; Smith 2003; Kostoff 2006; Estany & Herrera 2016. For the literature on innovation in other areas, see e.g., Shavinina 2003 (scientific innovation is dealt with in chapters by Nickles, Holton, Shavinina and Dasgupta).

an inexact, prescientific concept, the *explicandum*, into a new exact concept, the *explicatum*.” (Carnap 1950, 3)¹³ The metaphilosophical literature on this and other, partly more demanding, versions of “conceptual engineering” is growing (for an overview, see Brun, 2016). I need not engage with this issue here, since my aim is modest: I will assemble four plausible and necessary conceptual points that should be grasped if we aim to study scientific innovations in philosophically and historically fruitful ways. My list of conceptual reflections is not intended to be complete, neither does it deviate far from ordinary notions (as Carnap (1950, 8) noted, one feature of a good *explicatum* is that it retains sufficient similarity to the *explicandum*—along with being exact, fruitful, and simple). In Section III of this paper, it will become clear that even from my moderate account a puzzle arises which promotes the view that we need to improve our understanding of scientific innovations in one way or another. How to do that will depend on our theoretical or practical interests. But first, I introduce the four necessary, and I hope convincing, points.

II.1. A NETWORK OF CONCEPTS

It is useful to situate the notion of innovation within a broader network of related concepts. Often, no clear distinction is drawn between the *invention* of, say, a new scientific concept, instrument, and so forth, and the innovation to which that invention may lead. Thus, Thomas Nickles (2003, 59) approvingly cites Daniel Dennett stating that innovation is “bringing new design into being”. Nickles (2003, 59) also says that there can be *failed* innovation: it does not spread through the scientific community. However, in economics and technology studies, where the concept of innovation originated, we distinguish between ‘invention’ and ‘innovation’ (see e.g., Schumpeter 1934, 88f; Smith 2003). Researchers in those fields often add a third stage: ‘diffusion’. What matters for our present purposes is that it is one thing to *invent* new artifacts, but quite another to figure out that such inventions are *useful* (Sternberg 2003), and so can be used, thereby changing markets and societies. Thus, innovation only comes about after invention; and full diffusion, only after innovation.¹⁴ Applying these ideas to scientific change, we should say that a novel method or concept that is not used by (sufficiently large parts of) a scientific community is not an example of innovation, but “merely” an invention—and ‘invention’ need carry no positive connotation.¹⁵

¹³ When Carnap originally introduced the methodology of explication for the case of probability, he pointed out that this concept possesses irreducibly different meanings that need to be isolated from one another in order to make the right one precise for a specific theoretical purpose—here, the development of an account of confirmation of hypotheses. For his immediate purpose, only the “logical” meaning of probability mattered, i.e., the concept of probability as it applies to *judgments* (not to events or frequencies thereof) which possess *evidence* for or against them.

¹⁴ One innovation might later be replaced by another (think of the horse wagon, the car, the hybrid car, the self-steering car, etc.). In between, an invention will have to follow innovation; but this does not break the basic order of each invention-to-innovation cycle: an artificial entity *A* is first an invention before it turns into an innovation. However, any invention that then follows innovation *A* must be a *different* (type of) artificial entity, *B*.

¹⁵ At the same time, this distinction between invention and innovation differs from that between descriptive and evaluative questions. It is one thing to say that a new concept was invented by a scientist,

Clearly, this point is already a rectification that may be seen as going beyond ordinary language (or, more precisely, the language of grant agencies). However, there is at least one argument for this clarification: it builds on the theoretical fruitfulness of this conceptual choice. We can now better describe past attempts to change aspects of research that did not succeed, or at least not immediately, in the marketplace of scientific ideas. Consider applying the method of experimentation to psychology. This was first proposed in the eighteenth century by a mostly forgotten German philosopher-scientist, Johann Gottlob Krüger (1756), and practiced by only a few individuals at that time. Scientists back then already studied phenomena such as visual thresholds, the blind spot, the moon illusion, or the temporal persistence of visual and tactile perception. However, many scientists did not view these as *psychological* experiments,¹⁶ while some of those who did failed to fully grasp certain necessary conditions for proper psychological experimentation (Sturm 2006). Psychology did not become an academic discipline until the 1820s, when the Prussian state introduced it for trainee school teachers (Gundlach 2005); even then, it remained largely non-experimental. The establishment of laboratories only began in 1879 in Leipzig, under the direction of Wilhelm Wundt, and later spread across Europe and North America. Thus, we can say that while the method of psychological experiment was *invented* in the eighteenth century, that invention was not widely recognized (partly because it was insufficiently refined at that time to be useful), and the methodological *innovation* of psychology thus had to wait until a century later. That, again, does not mean that psychological experimentation was accepted everywhere instantly straight after 1879: the process of diffusion took several decades. It only means that at that point it had fully begun to enter the marketplace of psychological methods—literally, with companies starting to build instruments like precision clocks or recording devices for the new psychology labs (Gundlach 2007). Furthermore, once experimentation had reached a majority of academic psychology institutions, it could no longer be characterized as innovative; this is why it makes sense to distinguish between innovation and diffusion. Other examples for distinguishing more clearly between invention and innovation (and then diffusion) can be given. For instance, major breakthroughs such as Frege’s revolution in logic, Semmelweis’ introduction of antiseptic measures in clinical practice, or Copernicus’ heliocentrism were all initially met with skepticism or neglected, but later changed their fields dramatically (Gillies 2008). Using ‘scientific innovation’ in these ways, as one element in a network of concepts, provides science studies with more clarity.

This point also turns scientific innovation into a (partially, though essentially) social affair. One might object to this by distinguishing between “real” and “alleged” innovation: perhaps Copernicus’ heliocentric model of the world or Frege’s new logic *were* “real” innovations, even though at first they were poorly received or rejected by the community. Meanwhile other developments, such as ether theories in the eighteenth and nineteenth centuries, did not stand the test of time, even though communities had thought them to be

and then began to become accepted communitywide, and another to judge such an invention-to-innovation cycle to be *progressive*. Even large communities may go astray when accepting an invention. Consider the widespread and long-term popularity of concepts such as ‘phlogiston’ or ‘ether’.

¹⁶ There were a few authors in the eighteenth century, such as Johann Nicolas Tetens, who already viewed themselves as psychologists and who experimented with the mind (albeit with simple tools, and without a plausible theory of psychological experimentation; see Sturm 2006) in approximately the same way as current psychologists do.

highly useful, and so should at best be described as “alleged innovation”. However, I want to resist this temptation by pointing out that such an objection against viewing innovation as social conflates the explanatory and the evaluative questions: if we wish to explain a novelty as innovation, it is conceptually preferable to stick to the model of stages outlined above. Thus, any innovation requires an invention, which may be the product of an individual mind; but to be genuine innovation, it must be useful, and also start to become used.

One evident consequence of understanding innovation in this way helps to avoid another conceptual confusion. As with innovation in the economy or in technology, an innovative change in science should be more than incremental (Nickles 2015); while it need not be revolutionary in the Kuhnian sense, involving gestalt switches or world changes.¹⁷ Accepting the method of experimentation as an innovation depended, as well as on its diffusion, on continued improvements of the method, on clarifying under what conditions and to which objects it could be applied, and also on the social and technological organization of research. This implies that the new method was not a mere addition to existing practices: it often reshaped these or made wholly new types of research possible. That is why we view it as innovative. Innovation in science changes the research system in non-incremental ways; otherwise it is not innovation, or not judged to be.

II.2. THE ENTITIES THAT MAY BE SAID TO BE TARGETS FOR SCIENTIFIC INNOVATION

The second point is that the language of innovation is applied to a number of different entities, including: concepts, problems, methods or rules of scientific reasoning, theories, goals of inquiry, and institutions. These are all possible bearers of the predicates “is innovative research” or “does not lead to a research innovation” (and similarly, to carry over the previous point, of the predicates: “is (not) an invention”; “is (not) widely diffused”). The statements of funding agencies and Kuhn’s language agree on this flexibility. One consequence of this is that innovation differs from discoveries; the latter being specific cognitive achievements or research outputs: the discovery of Uranus, of two types of electricity, of nuclear fission, of Martin Heidegger’s “Black Booklets” (bad news!), or the effects of external temperature as well as of alcohol consumption on reasoning about the trolley problem (being in a cold environment allegedly enhances utilitarian responses, but so does drinking alcohol ... at least according to certain studies).¹⁸ Discoveries are a primary *goal* or *prod-*

¹⁷ As one anonymous referee has noted, Kuhn and later Kuhn scholars toned down the notion of revolution, so that what appear to be radical revolutions only seem so because our history of science is not sufficiently fined-grained (Kitcher 1993); or there could be gradual revolutions (Anderson, Barker & Chen 2006). I cannot enter into this here; suffice it to say that, as I explain in Section II.2, many different items can be subject to innovation—and perhaps if several (e.g., theories, instruments, and reasoning standards) are changed *at once*, we have a clear case of non-incremental change in science.

¹⁸ Nakamura et al. 2014; Duke & Bègue 2015. Both studies indicate that decreased empathy predicts utilitarian choices better than increased deliberation. So, if you are sitting in a cold room with lots of vodka in you, it seems almost unavoidable that you will decide on trolley paths like a true utilitarian would. Yet, one needs to add important qualifications here: presumably, if you are in a cold environment, your feeling cold leads to less empathy, while if you are drunk, you will feel warmer than you actually are, and then by virtue of the former correlation, you should be more empathetic. I suggest cognitive scientists should consider this important complexity in future research proposals.

uct of inquiry. In contrast, innovation is often the *means* through which to achieve new research outputs.¹⁹ Of course, a discovery can become part of the means by which later scientists innovate research; but for this, they usually need to appropriately combine the discovery—say, of a chemical element, an organism, a physical process, and so on—with other elements, such as methods, instruments, questions, concepts, models, goals, or even institutions.

A related point is that ‘innovation’, unlike ‘discovery’, is not an achievement term (Nickles 2003, 59). If you have discovered some process, entity, or fact, there is no way of undiscovering it. An exception to this may be the notorious case of Pluto. In 2006, after heated debate, the International Astronomical Union (IAU) voted that Pluto would no longer be a planet but, adding injury to injustice, should be reclassified as a “dwarf planet” (which was only partly redressed when being “plutoed” became the American word of the year in 2006). It would certainly, however, be an overreaction to say that this case undermines the basic intuition that the discovery of X is an epistemic achievement. The object called ‘Pluto’ was discovered in 1930; what the IAU did was to create a new concept and reclassify that same object between those of planet and small Solar System body (Dick 2006). Innovation, however, because it forms part of the cycles via which the sciences renovate the intellectual, practical, and institutional framework of research, can indeed be undone. How? Well, while a certain innovation is useful in one setting or context, it may lose its utility under different circumstances—when a new type of instrument appears, such as the advanced confocal laser microscope that may soon replace traditional methods in dermatopathology due to its enhanced speed and accuracy of diagnosis, and the help it offers in the treatment of skin cancer (Gareau et al. 2017). Furthermore, an innovative updating of instruments may become judged to be merely perceived rather than real. For instance, among neuroscientists, it is often claimed that positron emission tomography (PET) and fMRIs improve on preceding methods, such as the classical and widely used electroencephalogram (EEG), for studying the neural processes that underlie or realize mental processes. However, the epistemic innovativeness of PET and fMRIs has been questioned. Perhaps expectations are too high; or perhaps we have learnt that each different tool has its relative strengths and limits concerning the different processes and questions we are interested in (Roesler 2005; Bösel 2007). Our judgments concerning innovations are typically more open than in genuine discovery.

¹⁹ Schumpeter (1934, 66) distinguished between product innovation as “the introduction of a new good ... or a new quality of a good” and process innovation as the “introduction of a new method of production”. One might think, accordingly, that discovery, being the primary goal and product of research, is a kind of product innovation. However, we have to distinguish between discovery as such and recognition of it. According to the first conceptual point made above, innovation follows invention, and ‘innovation’ refers to the very first stages of the dissemination of an invention into a market. Therefore, applying Schumpeter’s distinction between product and process innovation to science correctly means that we would, strictly speaking, refer to the early stages of the recognition of a discovery, not to the discovery itself. Of course, some sociologists of science would object. Thus, Schaffer (1994) claims that we cannot distinguish between the two: the communitywide recognition of a discovery is all there is to discovery. But, apart from the fact that this is a *non-sequitur*, it seems to me that Schaffer’s view does not promote a more fine-grained analysis of different stages of innovation in science.

II.3. SCIENTIFIC INNOVATIONS ARE FREQUENTLY CONTEXT-RELATIVE

A third point follows closely upon the last remark: inventions and innovations are frequently context-relative (cf. Nickles 2003). What counts as innovation in research in one context may not count as it in another. When the ERC requires a proposal to be innovative, this is not in an absolute sense but only relative to the researcher's field. Along these lines, there were almost no departments of history of science in Germany before the early 1990s, so when the MPG founded the MPI for History of Science in 1994, it was an innovation for the German academic system. This was despite the fact that in the USA, the history of science was already better organized and widespread at that time. This is also what we see with the innovative introduction of experiment and measurement into psychology in the nineteenth century, although these methods had been used in physics throughout the early modern period.

We must not, however, overstate the point about context-relativity. While much scientific innovation is context-relative, some is not. When measurement or experiment was used for the first time in science, the innovation was absolute. When *simulations* using computers were introduced into science within the Manhattan Project to calculate the chain reaction process in nuclear fission by means of Monte Carlo methods, this too was absolute innovation. When Herbert Simon and Allen Newell then introduced computer simulations into cognitive science, that was clearly a context-relative change. The shift from one field to the other took decades because computer programs often did not function as planned, and moreover a generation of psychologists had to be trained to use computers in their daily routines (Gigerenzer & Sturm 2007): another example of an invention-to-innovation-to-diffusion process.

II.4. SCIENTIFIC INNOVATION REQUIRES RATIONAL DELIBERATION

A fourth point is crucial in paving the way for discussion of the explanatory, evaluative, and science policy questions introduced at the outset. It is this: innovation does not just happen, neither is it unavoidable. It is neither the result of mere chance, nor necessary in a strong sense. Nickles (2003) steers a middle course between these extremes by claiming that innovation is the outcome of an adaptive process of blind variation (BV) plus selective retention (SR). In addition to "blind" variation, Nickles also speaks (2003, 56) of "undirected" or "random" processes. However, he is careful to point out that BV and SR can be, and often are, "directed", due to constraints resulting from previous steps in evolutionary development. Biological species do not evolve in a purely random way, since variation is always a variation of a type that existed before. BV+SR can therefore lead to a certain directedness and can "have positive methodological significance" (Nickles 2003, 61).

However, is BV+SR the *only* way to understand processes of innovation? While it applies to biological innovations, with respect to scientific innovations, "blindness" is too strong a concept. It helps here to introduce a related but different metaphor: let us consider *short-sightedness*. To remedy the effects of short-sightedness, inventors developed glasses and later on contact lenses. Since these tools did their job, they became more widely used. Because not all eyes are the same, we have to interact with eye doctors and opticians to find sufficiently good optical tools through visual tests and measurements under varying conditions, the trying out of new types of lenses, and so on. Opticians, as well as their clients,

have a pretty good idea of what the aim is. They all engage in the process because they are aware of the deficiencies of short-sightedness, and have some causal knowledge of how to deal with them. Overcoming visual deficiencies thus involves a good deal of *conscious, deliberate means–ends reasoning* rather than BV. Similarly, scientific innovation can be the result of cognitive efforts and decisions because, as Nickles (2003, 63) accepts, innovations involve not just any kind of inventions whatsoever, but only the *useful* ones: processes of scientific innovation are *goal-directed*; they start not with BV, but with a kind of instrumental reasoning. So, what is it that constrains and guides the process? Certainly existing knowledge about reality, the research goals in relevant fields, and technologies capable of leading to new knowledge that leads us closer to the given research goals. However, in interesting cases where we really wish to move beyond what Kuhn (1977) called “convergent thinking”, it matters that we also possess reflexive or meta-knowledge: knowledge about *gaps* in existing knowledge, about *why they are there*, and about *what is preventing us from overcoming them*. We must grasp the limits of methods or technologies, because only then can we recognize what we can and cannot come to know, and we must also be ready to revise the goals that have so far guided our research. When we are promised scientific innovation, we should be presented with an invention that can become useful for making real progress in science. This meta-knowledge is not well described when we speak of BV+SR, since it requires a high level of informed and comprehensive or systematic thinking. Science is—at least it often is—highly systematic in its aspirations (Hoyningen-Huene 2013), in that it develops maps, as it were, of its own possible futures, including uncharted territories. Innovative research does not merely mechanically fill in gaps or make discoveries (no matter how interesting), but aims to understand the present state of the art in order to overcome stubborn obstacles to progress. The aim is to overcome some short-sightedness which plagues science, *purposefully*.

Another point that supports the idea that scientific innovation involves instances of deliberate means–ends reasoning is that innovation typically involves *violations* of existing rules of science. Each science is constituted as a set of rules that manifest themselves in research practices. When scientists propose an innovation, they mean to change (some of) those rules and to replace them by something new which they think might solve problems better. Rationality can only be ascribed to animals whose behavior is not merely regular, but also *rule-guided*.²⁰ Rule-guidance is only possible in creatures who can be aware that they are acting in accordance with a rule, that there can be better or worse reasons for both the action and the rule, and that one may therefore have reasons to violate or revise rules. This does not yet say anything about the nature and standards of reasons and rules; but it retains a minimal sense of the rationality of scientific inventions and innovation.²¹ If you doubt this, you should also doubt that anyone should submit proposals that claim to be innovative, or that funding agencies should assess scientists using this terminology.

²⁰ I am here building on Bennett’s (1964) analysis of a basic aspect of rationality which he illustrates through the distance between the quasi-language of bees and human linguistic behavior.

²¹ I use the term ‘minimal rationality’ here without committing myself to a specific set of *constitutive* rules of reasoning, though I agree with Cherniak (1986) that there must be some such rules.

II.5. SUMMARY

So these are my considerations in response to the conceptual question. (1) We should distinguish between the stages of invention, innovation, and diffusion. Scientific innovation requires inventions that are useful for changing scientific research practices or systems in non-incremental ways. (2) Unlike ‘discovery’, ‘innovation’ is not an achievement term: it applies to elements that make non-incremental, useful changes possible; but it does not necessarily itself consist of specific research results. (3) Innovation is often, though not always, context-relative. (4) Innovation is often developed purposefully and implies violations or revisions of previously established rules of science. In this sense, innovation presupposes (at least minimal) rationality.

I do not claim that these points fully define ‘scientific innovation’, if only because the term refers to a phenomenon without a clear essence. Furthermore, I do not maintain that the account provides a full explication in Carnap’s sense: while it delivers on the desiderata of sufficient similarity to the currently used meaning of ‘scientific innovation’, of fruitfulness (e.g., for understanding certain aspects of the history of science better), and perhaps also of simplicity, I do not maintain that the account is as exact as it might be. The concept studied here has fuzzy boundaries and moreover it results from actions, social rules, and interpretations concerning that concept which might change. Notwithstanding this, the account reflects how the concept is used and prevents some misuses by distinguishing it from related concepts such as invention, discovery, or revolution in science.

III. *A dilemma: Can scientific innovations be rationally explained?*

As we have now seen, processes of scientific invention and innovation typically involve a certain amount of deliberation and reasoning. Consider, in light of this deceptively unproblematic claim, the three further fundamental questions introduced in Section I: theoretical as well as practical questions concerning scientific innovation. They can now be restated as follows:

The *explanatory* question: How should we rationally explain the emergence of scientific innovation?

The *evaluative* question: How should we rationally evaluate claims of innovativeness?

The *science policy* question: How can we rationally foster innovative research projects?

That is, by what kind of deliberations, considerations, or reasons can innovation be explained, evaluated, and fostered? These are obviously complex and thorny issues. Answers will vary substantively depending on the specific cases or (real or alleged) scientific innovation, on the type of entities that are being innovated (theories, methods, instruments, goals, or institutions), on the degree of innovativeness (context-relative or absolute), and also on the validity of claims of innovativeness. These complexities mean that I cannot provide a full answer here; indeed, even building a conceptual scheme that categorizes them all is a task that has yet to be carried out. However, I wish to show that we are facing an interesting dilemma that needs to be clarified before any of the three questions can be addressed.

III.1. PRELIMINARY STATEMENT OF THE DILEMMA

The dilemma can be stated with the following two propositions:

- (1) Scientific innovation can be rationally explained.
- (2) No existing account of rationality explains scientific innovation.

This problem concerns the *explanatory* question. For reasons of space, I will mostly direct my remarks in what follows to this problem. *Mutatis mutandis*, the problem can be restated for the evaluative and the science policy problems as well.²² I will first argue that there are strong reasons to hold both (1) and (2). But there is an obvious tension between them, and we should think about how to respond to it.

Two main premises speak in favor of proposition (1). First, there should be ways of explaining how scientific innovation comes about. Innovation is not mysterious: we should not accept explanations that refer to supernatural influences (Nickles, 2003, p. 54). Second, however, it is not sufficient merely to explain causally how an innovative theory or method came about. We also need to understand why it was (viewed as) a *good* or *reasonable* move to invent or introduce some new element of research and then to disseminate it across a scientific community. Theories, concepts, instruments, models, methods, and institutions are not like Humean beliefs and passions, which we cannot avoid having once we are in a certain environment or situation. Scientists make deliberate choices. Importantly, however, those choices are made under conditions of a greater or lesser degree of uncertainty. When scientists propose that their next project is innovative in its methodology, questions, or goals, they are placing a bet; not a totally arbitrary bet, but one based on informed guesswork, or preliminary reasons. What is the nature of such reasons? Historians who have to clarify whether, say, Johann Gottlob Krüger made a rational proposal with the potential to lead to innovation in psychology, cannot avoid the problem by saying that issues of rational assessment are not their business. If they did that, they would not be able to explain genuine innovation *as* innovation.

III.2. WHAT ACCOUNTS OF RATIONALITY MIGHT EXPLAIN SCIENTIFIC INNOVATION?

Following on from the previous point, one therefore needs an appropriate account of rationality to address the explanatory question (and, *mutatis mutandis*, the evaluative and the science policy questions as well). This brings us to proposition (2). The reasons for this claim are more difficult to provide, since the field of rationality is highly fragmented: there are divisions between descriptive and normative approaches (and different normative theories); as well as instrumental vs. non-instrumental, universal vs. context-relative, and procedural vs. substantive notions of what rationality is (Mele & Rawlings 2004; Hanna 2006, xvi-xix). For my present purpose, I will focus on two major accounts: (a) the so-called

²² Thus, for the evaluative problem, the dilemma would be between: (1-E) scientific innovation can be evaluated on the basis of rational considerations; and (2-E) no existing account of rationality can justify the evaluation of scientific innovation. For the science policy problem, the dilemma would be between: (1-S) scientific innovation can be fostered by means of rational considerations; and (2-S) no existing account of rationality can help foster scientific innovations. Some important changes would have to be made, as we are changing from descriptive to normative domains, and rational considerations in each of them can be *ex ante* or *ex post*. I will leave this for another occasion.

“standard” account of rationality (Stein 1996; SAR for short), whose historical roots can be found in many traditions of epistemology and philosophy of science, but also in statistics and economics. It invokes strictly universal or formal and optimizing norms of logic, probability and decision theory. Opposed to it is (b) the “bounded” account of rationality (BAR), according to which good reasoning has to be explained and evaluated relative to the contents and contexts of reasoning tasks; and which refers to so-called “fast and frugal heuristics” (FFHs; e.g., Simon 1957; Goldstein & Gigerenzer 2002; Todd & Gigerenzer 2000). Both accounts are used in the social and cognitive sciences to explain judgment and decision making; and as normative yardsticks as well. What makes them particularly suited to the present discussion is that they are connected to different levels of aspiration for rationality: When it comes to the rationality of innovation, should we optimize or settle for less? What is the reasonable thing to aim for? While this looks like a purely normative question, it applies to the explanatory question as well: insofar as an explanation is a rational one, we can ask what level of aspiration was in fact adopted, and what rules or norms scientists actually used when working towards innovation.

III.3. THE STANDARD ACCOUNT OF RATIONALITY AND ITS PROBLEMS

Let us consider the SAR first. According to this conception, one has to answer epistemic questions as an *ideal* reasoner would if equipped with infinite time, unlimited reasoning abilities, and complete information: providing an *optimal* solution. Similarly, the point of rational choice theory is not merely to formalize requirements for consistent decisions, but also to combine preferences with probabilities such that we *maximize* expected utility. Of course, we should not burden the SAR with demanding a *guarantee* or *certainty* that the results will be achieved. We “merely” demand that it is probable that they will arise. We make decisions with associated *risk*. Beliefs we hold are risky because we base them on objective probabilities: there is no guarantee we will always arrive at the truth, but at least we are not fumbling in the dark.

Now consider a past funding proposal. The scientists submitted it with the expectation of a fruitful course of *future* research. How could rational choice theory be applied to the explanation of such a proposal? Obviously, one would have to develop (as an applicant) a proposal that maximizes expected utilities. That is, one would have to weigh and calculate all relevant probabilities—e.g., of discoveries to be expected—and utilities—e.g., the usefulness of those discoveries for further scientific progress, for the development of new scientific instruments or other technologies, etc. However, this idea of applying theories of probability and rational choice is highly unrealistic: scientists are not in a position to assess a proposed innovation in this way. One major reason for this lies in the difficulty of assigning probabilities here *at all*. Scientific innovations are rare, complex, and highly diverse. We cannot use standard statistics or frequencies of past success rates to ascertain the probability of a claim that an invention will turn out to be innovative being true. Not only do appropriate statistics not exist, it seems impossible to produce them. The same applies to the view that probabilities are subjective degrees of belief, to be determined by a willingness to bet.²³ Thus, the decision to make a proposal promising innovation was not risky but radically *uncertain*.

²³ For more on plausible theories of probability, see Gillies 2000.

A similar claim can be made when we turn, for a brief moment, from explanatory to evaluative considerations. The committee evaluating a funding application is, at the time of their decision, in a similar situation: the research has not yet been conducted. Committee members have to take a decision concerning a highly uncertain future. When the ERC (as quoted in Section I) demands that applicants present “high-risk, high-gains” proposals, it does not understand what it is asking for. There is even a trilemma here. First, by demanding high-risk ideas, the ERC assumes that scientists can make probability estimates, which they cannot. Second, if the ERC really intended to demand *uncertain* innovation projects, it should be made clear that this may well conflict with the additional demand that the project be “feasible”. If you want researchers to deliver a useful invention that is, at the moment the decision is to be taken, uncertain, then you should not demand feasibility: just give them a lot of money and time and freedom from, say, administrative work. That, of course, seems unreasonable too. The best that the demand for feasibility can mean is that researchers have already shown, for instance in pilot studies, that their research will be firmly rooted in previous knowledge, as well as in accepted theories, methods and instruments, and that their proposed “innovation” somehow carries a promise to work well. Here again we are relying on Kuhn’s (1977) demand that we balance “convergent” and “divergent” thinking. However, thirdly, such a proposal can hardly be said to be *highly* innovative, in the sense of leading to non-incremental changes. It should be clear by now why many academics have the suspicion (which they admit to in private conversation) that the language of innovation is often bombastic. A lot of good or excellent research can surely result from, say, ERC projects; but genuine scientific innovation may well be a qualitatively different thing.

Thus, a rational explanation of the development of a new theory or method, or a novel institution, cannot imply that the proponent somehow justified its adequacy in a way that would follow the SAR rules. Furthermore, if we *now* know that a past choice for a certain theory, framework, method, or instrument led to scientific innovation, this is because we have judged from its successes; which were unpredictable when the choice was made. A rational explanation of past events must take into account what resources or rules of good reasoning agents had at a given time, and no more. We should be reluctant to use our present knowledge for a rational explanation of the choices made by agents in the past, even if we might *now* be able to model such choices using current probability and rational choice theory. If we do such things, we will commit hindsight fallacy.

Nickles therefore rightly criticizes the attempts of grant agencies to apply approximations or surrogates of “logics of confirmation” to proposals, thereby ignoring the fact that such “logics” are backward-looking, not forward-looking normative theories of justification:

“In frontier research contexts, one will rarely have a partition of states of nature and their probabilities, and one is likely to be unclear about the ultimate goal and hence the preference ranking and utilities.” (Nickles 2016, 35)

There is another crucial point against using the SAR. When considering the conceptual explication of innovation, I argued that innovation must be distinguished from invention. Only inventions that are useful in one way or another should be considered as possibly leading to innovation, and only if they spread are they truly innovating. So, the *explanan-*

dum is not only the emergence of a novel element of research—a theory, instrument, or the like—but how it spreads through and leads to innovation in a scientific community. A lot of background information about inventions—concepts, theories, instruments, etc.—as well as about the modes and actual processes of reasoning or arguing for the novelty and usefulness of inventions will be required. This information is not given by the SAR as such. So once again, the SAR cannot deliver rational explanations of scientific innovation.

III.4. THE BOUNDED ACCOUNT OF RATIONALITY AND ITS PROBLEMS

What then can we say about the BAR? This notion of rationality uses FFHs as its rules—such as “tit for tat” (i.e., cooperate first, then imitate your partner’s last reaction; Axelrod, 1984), the “recognition heuristic” (when you have two options and know nothing but the name of one of them, assume that this option is what you are looking for; Goldstein & Gigerenzer 2002), imitate the majority (Boyd & Richerson 2005); to name but a few. These have been shown not only to be what we actually use, but also to be efficacious for many real-life problems, such as whom to marry, which job to take, or how to buy a new TV set; all of which are characterized by uncertainty and computational intractability. Based on a limited number of cues, FFHs use simple algorithms which, after a finite sequence of search steps, without weighing cues or calculating probabilities and utilities, deliver determinate answers. They work well, sometimes even better than optimizing rules, whenever there is a suitable and reliable connection between the reasoner’s mind and environment (Gigerenzer & Sturm 2012).²⁴ It is also important that FFHs, and together with them the BAR, avoid the commitment of the SAR to *optimal* procedures and solutions, opting for Simon’s satisficing instead. Finally, since the BAR is rooted in empirical research into human reasoning, it might seem reasonable to claim that this is what we should opt for when rationally explaining the emergence of scientific innovation.

Once again, however, there are problems here. First, proponents of the BAR themselves have emphasized limits to the applicability of the account:

“Some higher-order processes, such as the creative processes involved in the development of scientific theories or the design of sophisticated artifacts, are most likely beyond the purview of fast and frugal heuristics.” (Todd & Gigerenzer 2000, 740)

One reason for this is the following. FFHs work well when there is an objective answer which is hard or costly or even impossible for individual reasoners to figure out using only internal computation. However, in the case of many if not most instances of scientific innovation, no such objective answer exists. We have ideas or tools with the potential to lead to genuine innovation, but whether disseminating these throughout a community will prove to be useful or effective cannot be determined by simple heuristics, no matter how well they serve us in our daily decisions. We can certainly bet on the advantages of, say, transferring experimental or simulation methods to a new field of research, because they

²⁴ It is important that the normative adequacy of FFHs requires that there is an objective answer that one can look up in a lexicon (say, in the case of city sizes) or determine in some other way. This in turn might involve norms of the SAR; see Sturm, 2019. Here, I ignore this complication, since I focus on FFHs as possible rational explanations.

previously performed well elsewhere; but the devil is in the details. For instance, experimenting with human subjects poses new challenges, since they function differently from physical systems. Again, applying models and rules of evolutionary biology to psychology has seemed promising to defenders of evolutionary psychology; yet, its success has been limited (Richardson 2007). We can try out new approaches, but we cannot claim to know in advance that they *will* lead to innovative research. Of course, one might bet that, because a research group has been responsible for innovation in the past, they will be again in the future (a variety of the “trust the expert” heuristic). But while it may seem reasonable to think that someone who was an innovator of X will be an innovator of Y, rather than someone who has been responsible for no innovation to date, there are counterexamples in both directions. Linus Pauling, winner of two Nobel Prizes, devoted much of the end of his career to defending the (then as well as now) implausible view that a regular intake of high doses of vitamin C prevents cancer. Again, the Indian mathematician Srinivasa Ramanujan, with no formal training in mathematics, arrived at Cambridge in 1913 and worked with Godfrey Hardy to prove numerous entirely novel results, such as partition algorithms or the Ramanujan prime. The history of science provides spectacular failures as well as surprises concerning “trust the expert”.

Note, however, that all these considerations mostly relate to the evaluative and the science policy questions: the skepticism concerns the issue of whether one can use FFHs to predict or help bring about creative insight. What about the explanatory question that is under discussion here? In this case we encounter a second set of problems for the BAR. To begin with, we must distinguish *external* and *internal* judgments concerning the rationality of heuristics. That a judgment or decision can be described as boundedly rational because it fits with a certain FFH does not show, by itself, that scientists involved in invention–innovation–diffusion processes actually employed that FFH. Perhaps there is a temptation to think that, given that the SAR cannot provide the basis for a rational explanation of processes of scientific innovation, something else must do so; hence, the BAR is our best option. However, such an argument can again lead to hindsight fallacies.

Let us assume, nonetheless, that in specific cases we can show that scientists have applied heuristics. This is indeed possible. Then another problem has to be addressed: What is the relevant environment? Is that environment stable enough to assume that the heuristic will work in such a way that we can view it as a *reason*? When solving a specific scientific problem, the environment is probably constituted of a conjunction of methods, standards, instruments, theories, and the scientific community. But when using a heuristic to contribute to rational explanations of invention–innovation–diffusion processes, the mind–environment relation cannot be assumed to be stable—in part because by *changing* the rules of the game, innovators *change* the environment. They demand, for instance, that an instrument be used that did not exist before (and whose reliability may be in doubt, at least for a certain time) in order to prove a claim which could not be proved independently of using precisely that instrument. Consider Galileo’s difficulties in convincing his critics that the telescope is a reliable instrument, or that it can be trusted to support his observational claims concerning Jupiter’s moons. This was a complicated situation which could only be resolved over time. It involved a number of strategies through which Galileo showed that one can trust the telescope, and which in due time changed the whole practice of astronomy (Kitcher 1993, ch. 6).

To sum up the discussion so far: propositions (1) and (2) are both highly plausible, but also incompatible with one another. Admittedly, I have not *proved* that (2) is true, given

that I only surveyed two major theories of rationality. But these are very widely used for explanatory (and other) purposes. So, the dilemma stands.

III.5. HOW TO ESCAPE THE DILEMMA

Now, how to deal with this puzzle? One may weaken (1) as follows:

(1*) (Only) *some* scientific innovation can be rationally explained.

Thus, maybe we should not think that all scientific innovation requires a rational explanation; maybe sometimes we can only give a purely causal explanation. This would fly in the face of the fourth conceptual point; but I am willing to grant that there are different cases. Still, the problem remains: even if one merely claims that some innovation can be rationally explained, we still need an account of rationality for these cases.

Another way to avoid the dilemma is by questioning the meaning of

(2) No existing account of rationality explains scientific innovation.

After all, this does not mean the same as:

(2*) No account of rationality can explain scientific innovation rationally.

A truly deep, and insuperable dilemma or incompatibility between (1) (or (1*)) and (2) only exists if we take (2) to mean the same as (2*); but we do not have to do that. But then another challenge arises: What do we have to change in, or add to, our understanding of rationality? What would make research innovation through some invention a rational affair? Whatever it is, it has to incorporate the following two points with which I will conclude.

First, a conflation of discovery and innovation could easily lead to the expectation that it is possible to rationally explain innovation in terms of rules of (or good or even ideal) *epistemic* reasoning. Such a conflation is inadequate (see Section II): innovation is to be understood in terms of the initial recognition of the usefulness of an invention, so it is better understood as the right means for achieving scientific goals, including the discovery of truths. But even if discovery could be sufficiently explained by rules of induction (e.g., Langley et al. 1987), innovation cannot. Being located between invention and diffusion, and being not only about novelty but also about usefulness for a scientific community, innovation is a process which draws together theoretical, practical, and social aspects of scientific rationality. A sophisticated yet not too demanding form of instrumental rationality is needed.

Second, research can be innovative in *every* aspect. We not only change our beliefs or assumptions about reality, but also for example our methods, instruments, goals, or institutions. Scientific innovation is an unavoidably historical and social process; and explaining its rationality is therefore constrained by the actual history of science. It is nonetheless a minimally rational enterprise in the sense that, when innovating, scientists try to improve the rules of the game. This means that they must deliberately revise or violate existing rules and then convince others to join them; only then does research become innovation. If this is so, one cannot answer the explanatory question without looking closely at how scientists

have reasoned in the past. We must explain any innovation in terms of what *they* took to be good or optimal rules of reasoning, such as Aristotle's syllogism and his theory of informal fallacies; later logical systems; the theories of the *artes inveniendi*, *conjectandi*, and *iudicandi* of early modern thinkers; or theories of probability, induction, abduction, and the heuristic principles of research; and so on. We thereby tie our explanatory understanding of past innovation to the state of knowledge, the methods, and the existing problems at a given time in a given area of science; but also to the gradual deliberative reasoning by which novelties become accepted or used. There is, in other words, no rational account of scientific innovation without a historicized understanding of scientific rationality.

This helps to avoid hindsight fallacy in explanations of past scientific innovation. Moreover, the point also has consequences for the evaluative and the science policy questions. We may be able to learn from the past insofar as past rules of scientific reasoning resemble those we adhere to today; but perhaps we can also learn from a closer look at how past rules were violated or revised as inventions began to innovate, until they became completely disseminated throughout a scientific community. At the same time, it should be clear that past rational steps in innovative science only limit the future to a certain extent.

IV. Conclusion

As shown, the popular but imprecise language of scientific innovation requires clarification. We should view innovation as one stage within a larger process from invention to diffusion; and accept that innovation is a consequence of those inventions that are recognized as useful for changing research practices or systems in non-incremental ways. Furthermore, unlike 'discovery', 'innovation' is not an achievement term; and it applies to elements that make possible, but do not by themselves establish or guarantee, correct research outputs. Innovation is often, though not always, context-relative; and importantly, we assume that an innovation is deliberately developed and accepted, given that it implies violations or revisions of established rules of science: it is not an outcome of blind chance. In this sense, innovation presupposes (at least minimal) rationality. However, what notion of rationality can help us to explain, evaluate, and foster scientific innovation is a thorny issue; and I am inclined to think that abstract or general theories of rationality will not resolve it. I have shown this for the case of rational explanations; similar problems will arise when we consider the evaluative and the science policy questions. All this should make us question the widespread talk of 'innovation'. Despite our clear interest in seeing innovation in all fields of research lead to the advancement of science, calls for research proposals and offers of funding should be formulated in more reflective ways, and perhaps avoid direct demands for innovation; and their resolution should not hinge on excessively optimistic claims concerning our ability to predict and steer the future direction of the scientific enterprise.

Acknowledgments

For comments and discussions, I am grateful to David Casacuberta, Anna Estany, Catherine Herfeld, Thomas Nickles, and two anonymous referees. Christopher Evans helped to improve the language of this paper, and gave several highly useful recommendations con-

cerning content too. This work was supported by the Spanish Ministry for the Economy, Industry and Competitiveness (MINECO) through the research project *Naturalism and the sciences of rationality: an integrated philosophy and history* (FFI2016-79923-P).

REFERENCES

- Anderson, H., Barker, P. & Chen, X. 2006. *The cognitive structure of scientific revolutions*. Cambridge: Cambridge University Press.
- Axelrod, R. 1984. *The evolution of cooperation*. New York: Basic Books.
- Bennett, J. 1964. *Rationality*. London: Routledge.
- Boyd, R. & Richerson, P. 2005. *The origin and evolution of cultures*. New York: Oxford University Press.
- Bösel, R. 2007. Brain imaging methods and the study of cognitive processes. In M. Ash & T. Sturm, eds., *Psychology's territories*, pp. 275-286. Mahwah/NJ: Erlbaum.
- Brun, G. 2016. Explication as a method of conceptual re-engineering. *Erkenntnis* 81: 1211-1241.
- Carnap, R. 1950. *Logical foundations of probability*. Chicago: University of Chicago Press.
- Cherniak, C. 1986. *Minimal rationality*. Cambridge: MIT Press.
- D'Agostino, F. 2010. *Naturalizing epistemology: Thomas Kuhn and the 'essential tension'*. London: Palgrave Macmillan.
- Dick, S. 2013. *Discovery and classification in astronomy*. Cambridge: Cambridge University Press.
- Duke, A. & Bègue, L. 2015. The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition* 135: 121-127.
- Estany, A. & Herrera, M. 2016. *Innovación en el saber teórico y práctico*. Londres: College Publications.
- Gareau, D., Krueger, J., Hawkes, J., Lish, S., Dietz, M., Guembe Mülberger, A., Mu, E., Stevenson, M., Lewin, J., Meehan, J. & Carucci, J. 2017. Line scanning, stage scanning confocal microscope (LSSSCM). *Bio-medical Optics Express* 8/8, 1 August 2017.
- Gigerenzer, G. & Sturm, T. 2007. Tools=theories=data? On some circular dynamics in cognitive science. In M. Ash & T. Sturm (Eds.), *Psychology's territories*, pp. 305-342. Mahwah, NJ: Erlbaum.
- Gigerenzer, G. & Sturm, T. 2012. How (far) can rationality be naturalized? *Synthese*, 187: 243-268.
- Gillies, D. 2000. *Probability*. London: Routledge.
- Gillies, D. 2008. *How should research be organized?* London: College Publications.
- Goldstein, D. & Gigerenzer. 2002. Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109: 75-90.
- Gundlach, Horst. 2005. Reine Psychologie, angewandte Psychologie und die Institutionalisierung der Psychologie. *Zeitschrift für Psychologie* 212: 183-199.
- Gundlach, Horst. 2007. What is a psychological instrument? In M. Ash & T. Sturm, eds., *Psychology's territories*, pp. 195-224. Mahwah, NJ: Lawrence Erlbaum.
- Hanna, R. 2006. *Rationality and logic*. Cambridge/MA: MIT Press.
- Hoyningen-Huene, P. 2013. *Systematicity: The nature of science*. Oxford: Oxford University Press.
- Kitcher, P. 1993. *The advancement of science*. Oxford: Oxford University Press.
- Kostoff, R. 2006. Systematic acceleration of radical discovery and innovation in science and technology. *Technological Forecasting & Social Change*, 73: 923-936.
- Krüger, J. G. 1756. *Versuch einer Experimental-Seelenlehre*. Halle: Carl Hermann Hemmerde.
- Kuhn, T. 1977. The essential tension: Tradition and innovation in scientific research. In Kuhn, T. *The essential tension*, pp. 225-239. Chicago: University of Chicago Press.
- Langley, P., Simon, H., Bradshaw, G. & Zytkow, J. 1987. *Scientific discovery*. Cambridge, MA: MIT Press.
- Mele, A. & Rawlings, P., eds. 2004. *The Oxford handbook of rationality*. Oxford: Oxford University Press.
- Merton, R. 1968. The Matthew effect in science. *Science*, 159 /3810: 56-63.

- Nakamura, H., Yto, Y., Honma, Y., Mori, T. & Kawaguchi, J. 2014. Cold-hearted or cool-headed: Physical coldness promotes utilitarian moral judgment. *Frontiers in Psychology*, 2 Oct. 2014, Art. 1086.
- Nickles, T. 2003. Evolutionary models of innovation and the Meno problem. In L. Shavinina, ed., *The international handbook on innovation*, pp. 54-78. Oxford: Elsevier.
- Nickles, T. 2013. Scientific revolutions. *Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/entries/scientific-revolutions/>
- Nickles, T. 2015. Heuristic appraisal at the frontier of research. In E. Ippoliti, ed., *Heuristic reasoning*, pp. 57-87. Heidelberg: Springer.
- Nickles, T. 2016. Fast and frugal heuristics appraisal at research frontiers. In E. Ippoliti, F. Sterpetti & T. Nickles, eds., *Models and inferences in science*, pp. 31-54. Heidelberg: Springer.
- Richardson, R. 2007. *Evolutionary psychology as maladapted psychology*. Cambridge/MA: MIT Press.
- Roesler, J. 2005. From single-channel recordings to brain-mapping devices: The impact of electroencephalography on experimental psychology. *History of Psychology*, 8: 95-117.
- Schaffer, S. 1994. Making up discovery. In Margaret Boden, ed., *Dimensions of creativity*, pp. 13-51. Cambridge/MA: MIT Press.
- Schumpeter, J. A. 1934. *The theory of economic development* (R. Opie, Transl.). New Brunswick & London: Transaction Publishers.
- Shavinina, L. ed., 2003. *The international handbook on innovation*. Oxford: Elsevier.
- Simon, H. 1957. *Models of man, social and rational*. New York: John Wiley.
- Smith, G. 2003. Towards a logic of innovation. In L. Shavinina, ed., *The international handbook on innovation*, pp. 347-365. Oxford: Elsevier.
- Stein, E. 1996. *Without good reason*. Oxford: Clarendon Press.
- Sturm, T. 2006. Is there a problem with mathematical psychology in the eighteenth century? *Journal of the History of the Behavioral Sciences*, 42: 353-377.
- Sturm, T. 2019. Formal versus bounded norms in the psychology of rationality: Toward a multilevel analysis of their relationship. *Philosophy of the Social Sciences*, 49: 190-209.
- Todd, P. & Gigerenzer, G. 2000. Precis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23: 727-780.

THOMAS STURM is an ICREA Research Professor in philosophy and history of science at the *Universitat Autònoma de Barcelona* (UAB). His work centers on Kant's philosophy, rationality, and the relations between philosophy and history of science. He is the author of *Kant und die Wissenschaften vom Menschen* (2009), and (with P. Erickson, J. Klein, L. Daston, R. Lemov & M. Gordin) *How Reason Almost Lost Its Mind: The Strange Career of Cold War Rationality* (2013). He has also published numerous papers in journals such as *Erkenntnis*, *Kant-Studien*, *Studies in History and Philosophy of Science*, *Synthese*, *Philosophical Psychology*, and the *Journal of the History of the Behavioral Sciences*.

ADDRESS: ICREA, Pg. Lluís Companys 23 (08010 Barcelona), Spain. Dept. de Filosofia, Universitat Autònoma de Barcelona (08193 Bellaterra-Barcelona), Spain. Email: Thomas.Sturm@uab.cat