8-10-2018 2:00 PM

A Practical and Practice-Sensitive Account of Science as Problem-Solving

Frédéric-Ismaël Banville
The University of Western Ontario

Supervisor
Sullivan, Jacqueline A.
The University of Western Ontario

Graduate Program in Philosophy
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

Abstract

Philosophers of science have recently begun to pay more attention to scientific practice, moving away from the discipline's focus on theories. The creation of the Society for Philosophy of Science in Practice in 2006, as well as the emergence of scholarship on experimental practice (e.g. Sullivan 2009; 2010; 2016) as well as on the tools scientists use to construct explanations and theories (e.g. Feest 2011) all point to a disciplinary shift towards a more practice-conscious philosophy of science. In addition, scholars are realizing the potential social relevance of philosophy of science and identifying the obstacles that stand in the way of realizing said potential (e.g. Douglas 2010). The aim of this dissertation is to propose a novel descriptive framework for philosophy of science, one that rests on the idea that scientific *practice* should be described as accurately as possible so that philosophical discussions of science rest on a solid descriptive base and can more easily be of use in engaging users and practitioners of science. The dissertation is organized along the stages of developing what I call the dynamic-iterative model of scientific practice. In chapter 1, I clarify the project's philosophical background and outline six conditions of adequacy on a descriptive model of science. These conditions allow me to evaluate the three descriptive models reviewed in chapter 2. Chapter 3 combines Nickles' problem-solving view of science (1981) with core insights from the heuristics and biases literature into a workable version of the dynamic-iterative model. I then apply the model to the case of place cell research in chapter 4 as a test of the model's usefulness and applicability. In chapter 5 I evaluate the model with respect to my six conditions and assess its strengths and weaknesses. The dynamic-iterative model is a flexible descriptive framework that not only allows for more complete philosophical analyses of cases than other current models, but also supports important practical applications.

Keywords

Philosophy of Science, Philosophy of Neuroscience, Scientific Practice, Problem-Solving, Spatial Learning, Neural Representation, Heuristics

Co-Authorship Statement

Part of this dissertation discusses research conducted in collaboration with Jessey Wright. Specifically, section 6.1.1 is based on conference presentations co-authored with Jessey Wright, who contributed research on teaching methods.

Acknowledgments

This project is the result of a very complex and dynamic process which involved too many people to count, let alone name. That said, I can identify a few without whom it simply could not have happened.

I first want to thank my supervisor, Dr. Jacqueline Sullivan for her support, guidance and, most importantly, her patience. My decision to join the philosophy department at the University of Western Ontario was directly influenced by Jackie's arrival, and it was through my interactions with her that my project took on its current shape. I also wish to thank Jessey Wright for providing me with feedback, advice and, crucially, for his role as a semi-willing interlocutor during countless lunchtime brainstorming sessions. I am also grateful to Danny Booth, Chris Shirreff, Dylan Gault, Guillaume Beaulac and Robert Foley for their support and advice, academic or otherwise. I am indebted to the members of the Lab Associate program, especially those who made it possible, including Jackie Sullivan, Robert Foley and Christopher Viger, as well as to the Brain and Mind Institute community, and Dr. Melvyn Goodale, who welcomed me in his lab.

I also wish to thank my thesis examination committee for their time and insight: Dr. Gillian Barker, Dr. Melvyn Goodale, Dr. Tim Kenyon, and Dr. Wayne Myrvold.

I cannot express how thankful I am to my partner, Sarah Warren. Sarah, you have witnessed every stage of this project, and I don't think I could have seen it through without your unending support and faith in me. I love you.

I also wish to thank my family. Their support was invaluable throughout this process.

Funding for parts of this research was provided by the Social Sciences and Humanities Research Council of Canada as well as by the Ontario Graduate Scholarship program.

Table of Contents

List of Tables

List of Figures

Chapter 1

1 Introduction: the turn to practice

Disciplinary shifts come in many forms and seldom occur in a coordinated manner. Philosophy of science, concerned almost exclusively with the logical structure of *theories* and their epistemic justification from the 1950's onwards, is becoming more interested in accounting for scientific *practice.* This shift in focus is reflected in the establishment of the *Society for Philosophy of Science in Practice* in 2006, the emergence of scholarship on experimental practice (Sullivan 2009; 2010), on the role of measuring devices for the advancement of science (Chang 2004), and analyses focusing on the tools that support the generation of explanations rather than the explanations themselves (see, e.g. Ankeny 2000 on descriptive models in biology, Feest 2011). While topics of interest have shifted, descriptive models of science have not yet caught up. Philosophers interested in practice have been developing novel analytical tools or borrowing them from the social sciences, but there is no general descriptive model that can attend to these features of science that are becoming more central to the discipline. The aim of this dissertation is thus to propose a novel descriptive framework for philosophy of science, one that rests on the idea that scientific *practice* should be described as accurately as possible so that philosophical discussions of science rest on a solid descriptive base. To motivate this project, I provide a clear statement of the intellectual tradition within which the model I propose fits. This tradition is associated with what some scholars have called the "Practice Turn" (Soler et al. 2014b; Woody 2014), which corresponds to the emergence of greater interest in the study of scientific *practice*.

1.1 Motivations: descriptive accuracy and engagement

According to Soler at al., while the idea of a "Practice Turn" suggests a clearly identifiable point of origin (either temporal or theoretical) as well as a certain coherence, *neither are occurring in the context of the Practice Turn* (Soler et al. 2014b, 2). It is more appropriate to speak of a set of "practice trends", inspired by a shared set of methodological criticisms directed at traditional approaches to science studies writ large (viz. including the history, sociology and philosophy of science). Similarly, Woody characterizes the Practice Turn as a set of interrelated approaches that share a family resemblance *through their rejection of abstraction either in relation to the characterization of science or of scientists* (Woody 2014, 123). Thus, the definition of "Practice Turn" I develop here is primarily built on its negative aspects. By getting clear on what these approaches have in common, I aim to tease out a list of methodological requirements for a theory of science to qualify as "practice-oriented", viz. qualify as a theory which takes into account the important aspects of scientific practice while avoiding the descriptive inaccuracy associated with highly abstract theories of science.

Interest in practice-oriented approaches in philosophy of science and neighboring disciplines is evident from the emergence and increasing importance of the *Society for Philosophy of Science in Practice* (hereafter SPSP) established in 2006. The SPSP's mission statement provides some insight into the aims and methodological commitments characteristic of practice-oriented approaches. The mission statement begins with the observation that philosophy of science as a discipline is characterized by a focus on the relationship between scientific theories and the world, often ignoring or disregarding scientific practice. On the other hand, sociological and historical studies of science and technology overemphasize practice at the detriment of attention to the world, sometimes going as far as willfully discounting it as a product of social construction. The issue here is that each discipline is focusing on a relevant part of the scientific process while ignoring another, equally important part of it. What the SPSP advocates is simply a philosophy of scientific practice that takes *all* these aspects into account – theories, practice, and how those relate to the world.

The SPSP defines "practice" as "organized or regulated activities aimed at the achievement of certain goals" ("SPSP" 2017). The epistemology of practice, then, is tasked with elucidating the kinds of activities required to generate knowledge (at least, on the reasonable assumption that science aims at generating knowledge). The SPSP also believes that many traditional debates in philosophy of science may benefit from being recast in terms of practice. The general approach they recommend is, first, a concern with not only the acquisition of knowledge, but also the *use* of said knowledge, both in terms of understanding how pre-existing knowledge is applied to practical contexts, but also how the intended use of this knowledge shapes its very development. Second, they recommend increased attention to the way human artifacts such as conceptual models and laboratory instruments mediate the relationships between theories and the world (a topic van Fraassen's constructive empiricism dealt with extensively, along with Latour and Woolgar's ethnographical studies). Finally, the mission statement advocates for increased interdisciplinary outreach.

The program put forward by the SPSP is a concerted effort to get at the source of the discipline's (perceived) lack of impact on scientific and social spheres. On this topic, Heather Douglas remarks that "[standard] philosophical positions have not proven adequate for the contexts [that involve science practitioners and users of science]" (Douglas 2010, 318). The reason for this is the insular nature of philosophy of science. This insularity is in part due to pressures to de-politicize the discipline in the tense political climate of the 1950's which was accompanied by a narrowing of the set of acceptable topics for the discipline's flagship journal *Philosophy of Science* (Douglas 2010, 320-321). This process stands in stark contrast to the by-laws of the Philosophy of Science Association (PSA), which initially stated that the association would be dedicated to furthering a diverse research agenda and emphasized the importance of socially relevant practical consequences (Douglas, 2010, 320), and is reflected in changes to the journal's content over the years (see Douglas 2010, 320 for a detailed list of the topics and authors). An example of this disciplinary narrowing is the introductory chapter to Feigl and Brodbeck's anthology (1953) in which Brodbeck argues that questions related to the social context and significance of science, the social role of scientists and any related ethical issues are not part of philosophy of science proper. Philosophy of science, on this view, is directed at epistemological questions about science: what explanation is, how theories are justified or confirmed, etc. (Douglas 2010, 321). Philosophy of science's low impact factor outside of its academic sphere is a result of the discipline being insular: its concerns do not match those of practitioners and users of science[1].

Scholarship on the Practice Turn suggests the movement was a direct reaction to perceived inadequacies of established approaches in science studies writ large[2], and that its criticisms were targeted at the traditional approach's lack of descriptive accuracy and relevance for the day-to-day practice of science. I take descriptive accuracy to be a necessary step towards my goal of bolstering the potential impact of philosophy of science outside of its disciplinary boundaries. Although it likely is not sufficient, descriptive accuracy will be a core concern as I develop my descriptive framework.

Given this uniformity in what is rejected, the fact the criticisms are made in the spirit of making science studies more descriptively accurate, and the fact that this failure in descriptive accuracy is coextensive with the difficulty of applying insights from philosophy of science outside its disciplinary sphere, I take it as given that understanding *what* is rejected in terms of research goals and methods for philosophy of science amounts to identifying those disciplinary concerns that led to philosophy of science's insularity. Further, it amounts to identifying the aspects of traditional philosophy of science that weakened the descriptive accuracy of its mainstream theories. The crucial piece needed to understand the motivations behind the Practice Turn is thus a clear statement of what it was a reaction to.

---

[1]Douglas' proposed solution is coherent with the methodological recommendations associated with the Practice Turn. Specifically, she argues that the best way out of the quandary posed by the discipline's insularity is to adopt an "on-the-ground" methodology, which she admits is nothing new (without citing specific examples). It seems reasonable to assume she is thinking of ethnographic work in the vein of Latour and Woolgar's *Laboratory Life* (1986). Where she innovates is that she encourages the actual engagement of philosophers in socially relevant contexts (e.g. policy circles). I believe it important to flag that my project stops short of providing a full account of the role of social values and social contexts for analyses of science, but as I demonstrate in later chapters, the model can easily accommodate these factors, at least in principle.

[2] Soler et al. (2014) assert that the Practice Turn occurred both in philosophy of science and in history and sociology of science. In what follows I focus on the philosophical side of things, as it can be fairly easily separated from the other disciplines, but the influence that shifts in history and sociology of science had on practice-oriented philosophy of science should not be understated. Ethnographical studies such as Latour and Woolgar's (1986) had a fair bit of influence on the development of philosophical approach

1.2 Sources of insularity

“Traditional” philosophy of science is understood as any account that takes theory choice and confirmation (or justification) of theories along with the explication of the logic that governs these processes to be the central concerns of philosophy of science, typical of the early-to-mid 20th century. This overlooks the nuances of these views as well as a number of divergent projects that were of interest at the time, but this picture, truncated as it is, will suffice for the purposes of understanding the context in which the Practice Turn occurred[3]. Note that both the research goals *and* the methods of traditional philosophy of science contribute to this problem with descriptive accuracy, since research goals partially determine which methods are used. I treat those in isolation but will draw attention to key interactions when appropriate. I also want to make very clear that I do not accept this picture of 20th century “mainstream” philosophy of science as historically or conceptually accurate. However, this is the understanding proponents of the Practice Turn had of the discipline when they formulated their criticisms, and thus it is the picture we need to understand if we are to properly understand the methodological shifts that informed practice-oriented approaches.

In this picture, Carnap and Popper act as representatives of the “traditional” approach. Hacking remarks that Carnap and Popper “disagreed on much, but only because they agreed on basics” (Hacking 1983, 3). First, both held natural science to be the highest form of human rationality. That is, both considered science to require strict adherence to rules of logic (although they disagreed on the status of inductive reasoning as Popper was an inductive skeptic). Second, both viewed observation as entirely separate from theory and, importantly, held a clear distinction between the context of discovery and the context of justification. Discounting the conditions in which discoveries are made as work for other disciplines (sociology or psychology), both Carnap and Popper saw the job of the philosopher of science as that of evaluating the *product* of scientific work

---

[3] If the picture of traditional philosophy of science I sketch here feels like a strawman, I can only point to its resemblance to the picture provided in scholarship on the Practice Turn, and stress that whether it is a fair representation of these theories is somewhat irrelevant since the Practice Turn was a reaction to *this* understanding of traditional philosophy of science.

(theories in particular); she is concerned with providing an account of the *justification* of scientific products and with providing a clear logic of confirmation and theory choice, both of which are seen as the central concerns of scientists (Hacking 1983, 6). In what follows I break down the traditional view of philosophy of science into three sets of claims: claims about the nature of science (its aims, its structure, etc.) (e.g. Popper 1959; Nagel 1961), claims about the nature of philosophy of science (its aims, its methods, etc.) (e.g. Feigl and Brodbeck 1953; see also Douglas 2010 for a short history of the evolution of philosophy of science) and claims about scientists (specifically, whether scientists fit a certain model, how they approach their work and how they *should* approach their work) (e.g. Popper 1962).

The first set of claims concerns the goals of science and its status as the archetypal rational human activity. Science is thought, on this view, to aim at uncovering the truth about the natural world, to provide an understanding of the natural world[4]. For instance, Popper holds that scientific theories aim at the truth (although we can never know that a theory is true, only that it has not been refuted yet). This view of science is echoed in folk understanding of science (Lombrozo et. al 2008), as well as in certain science education initiatives (see for instance "Science Aims to Explain and Understand" 2016, a page developed in the context of the Understanding Science initiative). Importantly, there is no indication that critics of the traditional view disagree that scientists pursue such goals. The issue is that such goals may be too far removed from the day-to-day practice of science to be motivationally relevant (Rouse 2002, 104, see also section 1.3 below). Traditional views take these goals as the central motivations of science and draw a line from these goals to the way scientists should work if they are to attain them. Suppose that the goal of science is to explain and understand. If science explains and understands the

---

[4]Here "understanding" is intended to capture explanatory understanding. This is distinct from what one might call "maker's knowledge" of the world, which is a kind of practical understanding that makes it possible to, say, replicate the outputs of a natural process. As an example of the unclear status of this kind of knowledge, Craver's doubts about the epistemic value of brain-computer interfaces and prosthetic models based on such interfaces are partly due to the distinction between explanatory knowledge and "maker's knowledge", where the latter is a lesser form of understanding that simply enables us to reproduce the function of a given system without having to understand how the system itself *actually* works, so long as we get the desired result (see Chirimuuta 2013 and Craver 2010).

world by devising theories, then the core mechanic of science has to do with theories. The notion of theory appears to be a prime candidate for a unit of analysis, an "end product" of the scientific process that can be assessed and thus the most pressing task for philosophers of science is to explicate how theory-choice functions and should function. It is unsurprising that experiment and observation are seen as secondary aspects on this view, only relevant to the extent that they provide supporting or disconfirming evidence for theories (Soler et al. 2014b; Hacking 1983). In fact, in at least one recent work explicitly presented as a teaching resource for philosophy of science, i.e. Godfrey-Smith's *Theory and Reality* (2003), discussions of conceptions of experimentation only meaningfully come up in chapters devoted to sociology of science. This is telling given the book's chronological structure (Godfrey-Smith 2003, xiii): concerns about experimentation only became part of the disciplinary landscape in the 1980's with the publication of Latour and Woolgar's ethnographical studies of laboratory practice (1979) and Shapin and Schaffer's study of the dispute between Hobbes and Boyle on the role of experimentation in science (1985).

This conception is accompanied by a corresponding view of the job of the philosopher of science, which is to provide an account of how theories are justified (either through confirmation or corroboration) and how scientists choose which theories they accept or reject. The general strategy is to identify the logical rules that govern the justification of theories. Carnap thought *induction* was an appropriate approach. One could make observations and build up to general statements on the basis of these individual observations (Hacking 1983, 3). On this view, scientific rationality is a matter of carefully supporting hypotheses and theories through observation until a theory is sufficiently well-supported and accepted. On the other hand, Popper rejected induction for much the same reasons as Hume, viz. that our psychological propensity for generalizing from experience provides no logical basis for inductive generalization. Popper holds that scientific rationality is a matter of *deduction*, of forming hypotheses from which we can deduce some empirical consequences that can then be tested (Hacking 1983, 4). The overall picture here is one where the philosopher of science imposes a logical structure onto science, essentially reconstructing it through the lens of whatever logical system she develops and is supposed to account for the way theories are justified and ultimately accepted or rejected[5].

Because of this very narrow focus, traditional approaches commonly draw a strict distinction between the context of *discovery* and the context of *justification*. Philosophy of science has historically been concerned with the latter: is the theory or hypothesis that results from the discovery sound? Is it reasonable, or well-supported? The circumstances of discovery, which include things such as who made the discovery, what social context it was made in, who funded the research, etc. are of no interest to philosophers (Hacking 1983, 6). History, and accompanying contextual factors, are not entirely ignored by traditional approaches, but used "only for purposes of chronology or anecdotal illustration, just as Kuhn said" (Hacking 1983, 6). This sharp separation between discovery and justification and its accompanying disregard for the details of history will turn out to be a common target for critics of traditional approaches.

The *a priori* approaches characteristic of traditional philosophy of science had a clear advantage, however: normativity comes for free. Because approaches built on *a priori* principles specify rules of reasoning and rules for evaluating evidence, the normative component of philosophy of science comes built-in with the descriptive component. If scientists do not follow the rules derived from the basic principles, then they are not doing what they *should* be doing. The view of scientists as idealized rational agents seems to be a background assumption in many theories, despite the rarity of explicit discussions of conceptions of scientists. In part, this lack of attention to scientists and what they *can* do suggests that this was not a pressing concern for philosophers of science. Popper's theory of scientific change is a notable exception. Scientific change happens through a two-step process of *conjecture* and *attempted refutation* that repeats endlessly (since Popper is an inductive skeptic, we can never know that a theory is true,

---

[5] Note that the idea that acceptance and rejection are the only attitudes scientists can have towards a theory is roundly rejected not only by proponents of the practice turn (see, e.g., Hacking 1983, 15) but also by philosophers who attempted to salvage the general spirit of traditional philosophy of science while attempting to develop more descriptively accurate accounts of science. Laudan, in particular, argues that there is a much broader spectrum of attitudes scientists adopt towards theories, including pursuing a theory, or entertaining a theory (Laudan 1981b, 144). I discuss this in more detail later.

only that it has not been refuted yet). The first step involves a scientist offering a conjecture about some aspect of the world that needs explaining. A good conjecture is one that enables novel predictions. The second step involves subjecting the hypothesis to critical tests to show that it is false. At this point, the process starts over again: regardless of whether the hypothesis was refuted, the scientist comes up with another conjecture, which is then tested, and so on. On this view, a good scientist exhibits two key characteristics. First, she can come up with imaginative, bold ideas, and original contributions. Second, she needs to be unwavering in her willingness to subject her ideas to rigorous testing, no matter how ground-breaking, or, in some cases, how attached she may have become to those ideas (see Godfrey-Smith 2003, 60–62 for a more detailed discussion). On some readings, this view imposes stringent demands on scientists, both in terms of their cognitive capacities and of their behavior (Thornton 2016; Andersen and Hepburn 2016, see also Mayo 1991 for a discussion of novelty and severity of tests).

This view of scientists as highly rational (to the point of idealization) is a common target for criticism. Woody (2014, 126) uses it as a contrast point between traditional and practice-oriented approaches, that is, as one of the main points of departure between traditional views and their critics. Bechtel and Richardson (2010, chap. 1) discuss how the traditional view of scientific rationality is descriptively impoverished because it requires scientists to reason in a way that is not realistic for human agents with limited computational resources (a point also made by Wimsatt 2007a, chap. 1)[6].

Overall, traditional philosophy of science is characterized by its construal of what the core concerns of science are (to provide explanations, to accept the best theories, to justify those theories), its construal of what the task of philosophy of science is (to provide a descriptive and normative account of the logical structure of explanation, of the justification of theories, of theory-choice) and a set of theoretical commitments that

[6] Importantly, this idealized view of scientists as well as its descriptive inadequacy is the cornerstone of the project developed in this dissertation insofar as the project itself rest on an accurate description of the cognitive capacities of practitioners of science.

follow from these (the discovery/justification distinction, the construal of scientists as ideal rational agents, the lesser importance of history). As discussed in 1.1, the motivation behind the practice turn was, in effect, the rejection of this general view of science and philosophy of science, largely because of its descriptive inaccuracy. In this section, I have provided a picture of what proponents of the Practice Turn are rejecting when they reject traditional approaches to science studies. In the next section I discuss some of the shared methodological prescriptions put forward to remedy the situation.

1.3 Core tenets: four shifts

Soler and colleagues view the practice turn as a collection of "practice trends" which share a family resemblance and are unified by what they reject (see Soler et al. 2014, 3). This family includes widely diversified approaches, from ethnographic studies of scientific practice on the side of sociology of science (e.g. Latour and Woolgar 1986) to Hacking's "new experimentalism" (1983) on the side of philosophy of science. Diverse as they may be, there is an identifiable common motivation behind such projects: traditional approaches developed overly idealized and truncated accounts of science which are inappropriate as descriptions of actual scientific practice (Soler et al. 2014a, 11). Beyond the obvious (inaccurate accounts of something should be rejected because they are inaccurate), an important reason to reject a descriptively inaccurate account is that poor descriptive accounts make poor foundations for normative accounts. Wimsatt harshly criticizes traditional models of scientific rationality (or, rather, traditional models of rationality as applied to the analysis of the practice of science) on the basis that not only are they overly idealized and inaccurate, but that trying to adhere to them might actually be counterproductive: "[their normative status] *hides how badly these idealizations perform as descriptions of our actual behavior and also give misleading advice about what we should do about the deviations* we do *detect"* (Wimsatt 2007, 17, emphasis in original). It is possible for a normative model to demand that certain corrective procedures be implemented, but that implementing such procedures could be actively detrimental to the process they are supposed to improve[7].

This section looks at what proponents of the Practice Turn came up with to rectify the situation. Because the Practice Turn itself is a complex phenomenon in terms of history and theoretical diversity, my strategy is to focus on broad commonalities. Woody's discussion of chemistry's periodic law includes a useful summary of the Practice Turn in terms of four interrelated shifts in focus, each being a "retreat from a certain form of abstraction" (Woody 2014, 123-125). This interpretation is coherent with Soler's and with the one I adopted, as Woody's idea of four shifts away from abstraction can be rephrased as a rejection of the abstraction characteristic of traditional approaches.

The first shift is one from conception to representation. Theories are construed as artifacts or representations as opposed to abstract conceptual objects characterized by their logical structure. It is not the case that theories are now devoid of any logical structure. Rather, this is a shift in how theories are studied by philosophers of science. The idea is akin to the relationship between propositions as abstract objects and the specific statements that express them (Woody 2014, 123). Theories are evaluated through the models they generate, their predictions, specific explanations and so on, as opposed to being evaluated on their logical structure or soundness alone. Woody's analysis of chemistry's periodic law (a theoretical entity) is an example of this shift. She argues that the predictive success of the periodic law was highly contingent on the representational tools used to express it (Woody 2014, 126-127). Specifically, the representational format of Mendeleev's tables was such that it made it very easy to predict the existence of new elements since "gaps" would appear in the table. Meyer's graphical representations, on the other hand, showed no such gaps. Neither representational format was particularly well-suited to making predictions about the *properties* of new elements, however, but because of social factors

---

[7] As an example, triage systems used by volunteer rescue workers rely on very small sets of cues to allow volunteers to assist in triaging injured patients effectively even though they might rely on cues that would be diagnostically sub-optimal in, say, a hospital setting. Moreover, requiring exhaustive medical knowledge from volunteer rescue workers or imposing a specific kind of corrective procedure that relies on such knowledge would end up hampering triaging efforts, and thus would have undesirable consequences.

borne out of the scientific community's goals, there was enough value attributed to the first kind of predictions to ensure the entrenchment of Mendeleev's tables in the discipline (Woody 2014, 137)[8].

The second shift is one from *a priori* to empirical methods, especially with regards to the normative aspect of philosophy of science. Normative principles are built bottom-up from an observation of scientific practice as opposed to being derived top-down from *a priori* principles. More generally, the Practice Turn replaces *a priori* analyses with empirical examinations of scientific reasoning as it happens in particular contexts (Woody 2014, 123). This is a major departure from the traditional approach, and it is accompanied by a number of new methods derived from the sociology and history of science. An example of this is found in recent work in the philosophy of neuroscience. For instance, Sullivan (2009) argues that paying attention to the basic features of experiments in neuroscience highlights serious problems for Bickle's "ruthless reductionism" (2003, 2006), the idea that statistically significant differences in behavioral performance which are attributed to genetic or pharmacological manipulations warrant the claim that the molecular changes that result from these manipulations (and thus, the interventions) *explain* the behavioral changes (hence Bickle's phrase "reduction-in-practice"). Sullivan's argument begins with the identification of two basic features of experimental practice in neuroscience[9]: 1) the *experimental design*, which includes general specifications on how to set up the experiment (materials and methods used), and 2) the *experimental protocol*, which includes step-by-step instructions for running the experiment. After showing that designs and protocols are not identical across laboratories (one may use a different drug than another, for instance), she argues that the model is not in good evidential standing with respect to actual experimental practice. In Bickle's case, it is because the differences in

[8] There are two things to note here. First, Woody takes her analysis to be representative of all four of the shifts associated with the Practice Turn. My use of her argument as an example of the first shift should not be taken to imply otherwise. Second, she highlights the fact that the view she develops is novel in the sense that it links theory with practice, while it is more common to see work focusing on experimental practice that eschews any focus on theory.

[9] Sullivan is primarily interested in behavioral and electrophysiological experiments in cellular and molecular cognition research on learning and memory, which is the field from which the purported evidence for the models she discusses is taken.

experimental designs and protocols across laboratories undermine the strength of his claim that experiments reduce cognitive and behavioral changes to molecular changes. At best, Sullivan argues, these constitute local reductions, with nowhere near the generality that Bickle takes them to have and he does not provide an account of how they could be generalized (Sullivan 2009, 522–523)[10]. Moreover, Sullivan remarks that scientists must fulfill conflicting requirements in the form of the reliability of the data production process and the (external) validity of the interpretive claims made on the basis of said data. Reliability requires that the data allow for the discrimination of competing hypotheses in the context of the laboratory, which imposes constraints on the data production process. External validity, on the other hand, requires that interpretive claims be true of the phenomenon as it occurs in the world, which requires the inclusion of additional factors in the experimental design. The result is that reliability pushes towards simplifying experimental design while external validity pulls towards increasing their complexity (Sullivan 2009, 535). In conjunction with the diversity of investigative aims found in scientific research these conflicting prescriptions give rise to a multiplicity of experimental protocols. While this multiplicity is not itself an issue for the discipline, it does undermine the global unificatory picture put forward by Bickle. This shows how attention to actual scientific practice can be crucial for philosophers of science.

The third shift concerns the way scientists are described in philosophical theories of science. The ideal rational scientist is replaced by a human practitioner, the characteristics of which can be drawn from various sources including empirical psychology, folk psychology, or any other relevant account of human cognitive capacities and skills. Woody's analysis of chemistry's periodic law "makes crucial reference to the general perceptual capacities of practitioners" (Woody 2014, 145), inasmuch as the entrenchment of Mendeleev's tables was in part due to the fact they were easier to use than Meyer's graphical representations (Woody 2014, 137, 145).

[10] Sullivan also criticizes Craver's idea of "mosaic unity of neuroscience" (2007), but her criticism of Bickle is sufficient as an illustration of this methodology.

I want to insist on the importance of this shift. The analogy between the practice of normal science and puzzle solving proposed by Kuhn serves as a good illustration of the way in which psychological factors matter. Rouse explicates the analogy as follows: during normal science, certain well-defined "puzzles" constitute the bulk of scientific research and are assumed to have unique and accessible solutions, much like traditional puzzles do (e.g., crosswords puzzles). An apparent weakness of the analogy – that "puzzles" are trivial and lack the intellectual significance we usually associate with science is in fact its most important insight. The apparent disanalogy actually captures the way the "big goals" of science are often too removed from day-to-day activities to be motivationally significant (Rouse 2002, 104). Similarly, it is not an unreasonable view that the very high-level theoretical concerns on which scholars like Carnap and Popper focused are not part of the day-to-day issues that scientists grapple with[11]. They probably do think about those but are more preoccupied with proximal goals. This is why their attitudes towards theories are more varied than simply "acceptance" and "rejection". The fact that Newton's laws do not apply in certain circumstances (extremely small scales, for instance) might justify their rejection by a perfectly rational scientist, but the fact is they are a reasonably good approximation that gets the job done most of the time. The mental stance of a scientist applying Newton's laws cannot be said to be one of straightforward acceptance, but it isn't one of rejection either. Unless one understands that scientists pursue smaller-scale goals than that of "understanding nature", these attitudes are difficult to understand. Similarly, the very idea that there is something like an identifiable "traditional" philosophy of science is not one I fully endorse, but I use it here as a reasonably good approximation because it allows me to formulate a workable characterization of the Practice Turn for the purpose of teasing out methodological recommendations that I can then use to formulate a descriptive model of scientific practice. Discussing scientific rationality, Hacking notes that "accepting and rejecting theories is a rather minor part of science" (Hacking 1983, 15) and agrees with Laudan

[11] This points to the importance of the psychological factors involved in science. This is a major concern both for some of the authors discussed later in the chapter and, crucially, for the model I develop in the dissertation. I discuss it in more detail below, but it seems relevant to flag it here, as the lack of interest for the psychology of scientists is another problem of the "traditional" framework.

that there are many cognitive stances scientists can take with respect to theories beyond these two. The pragmatic use of a theory is one such example. This shift is crucial, as the problem-solving approach that grounds the model I develop in later chapters is a direct consequence of taking seriously the cognitive limitations and characteristics of human agents. Instead of idealizing scientists as perfectly rational, practice-oriented scholars such as Nickles, Bechtel, Richardson and Wimsatt all rely on empirical psychology and observation of in-context scientific practice to develop their accounts of how science functions. Similarly, scholars such as Sullivan and Woody rely on naturalistic descriptions of scientific practice which take psychological and sociological factors seriously. Sullivan's point with regards to the conflicting desiderata of reliability and external validity is essentially a point about the psychological and social factors that motivate scientists (2009). Woody's account of the acceptance of Mendeleev's tables also relies on an identification of psychological and sociological factors, viz. Mendeleev's tables were easier to use and more in line with the goals of chemistry's community.

The fourth shift is one from the knowing (individual) subject to social epistemology, recognizing the intrinsically social nature of science and thus shifting the discipline's perspective from that of the individual to that of scientific communities. Woody's analysis of the periodic law rests in large part on identifying the goals of the communities that used the various representations of periodicity, and these goals provide part of the explanation for the entrenchment of Mendeleev's tables (Woody 2014, 145). This stands in stark contrast to the kind of analysis that could be proposed by an imaginary proponent of the traditional approach, which would likely not include social factors in an explanation of the adoption of the periodic table as those are seen as external to philosophy of science.

With this understanding of the Practice Turn in hand, I can now explain how my project fits within that tradition and provide a clear statement of how I proceed.

1.4 The need for a new descriptive framework

The Practice Turn can be understood as a shift away from methods that rest on a particular type of abstraction, viz. idealized views of science and scientists. Moving away from such abstractions also motivates a shift in methodology, as the understanding of what the job of philosophers of science is shifts along with a better understanding of how science actually functions, how practice unfolds. To go back to the SPSP's mission statement, focusing on scientific *practice* amounts to focusing on the activities that scientists engage in to meet certain goals. Depending on what these goals turn out to be, scientists will engage in different activities, and it is reasonable to assume that different activities correspond to different contexts of practice, each with its own epistemic norms and standards. The four shifts discussed above form a loosely cohesive set of methodological prescriptions and I take them to constitute the core of practice-oriented approaches in general. Each of them is a specific case of the overarching rejection of abstraction-as-idealization and in that sense the Practice Turn is a shift in *methods*. One way to think about this methodological shift is as a shift towards *naturalistic* methods. Such methods focus on providing a faithful description of scientific practice, one grounded in empirical observations of said practice and any related factors (psychological, historical, sociological, etc.).

Such approaches seem to stand a much better chance to be descriptively accurate, but this accuracy does not come for free. Indeed, practice-oriented approaches deal with some challenges which are direct consequences of their methodological commitments. The first is that because of their focus on case studies, they run the risk of getting stuck at the level of particularity and may struggle to provide insight beyond the scope of these cases (Woody 2014, 125). This worry is reinforced by the fact that the program outlined in the SPSP's mission statement seems to suggest the kind of diversity of epistemic standards and, perhaps, descriptive accounts, discussed in the previous paragraph. This is not a problem in and of itself, but having an overarching or systematic model of science (while not strictly speaking necessary), would be beneficial not only for the discipline itself because it would provide a way to bridge a plurality of case-level analyses, but also for its potential for engagement outside of academic circles and in interdisciplinary contexts, because the resulting analyses of scientific practice would be descriptively accurate and thus would serve as a solid basis for evaluative and normative claim as well as direct collaborations. In both cases, I am attempting to formulate a description of scientific practice that would allow philosophers of science interested in practice as well as users and practitioners of science to "speak the same language".

The second challenge is a consequence of the abandonment of idealized views of science. Because these come with built-in normativity, those normative principles need to be secured from a different source. The preoccupation with descriptive accuracy that characterizes practice-oriented approaches comes with the risk of being purely descriptive (Woody 2014, 125). I do not mean to imply that practice-oriented philosophers approach science without analytical tools at all. My point is that the emphasis on descriptive accuracy can, potentially, result in lesser attention to normative matters, and while I take accurate description to be a prerequisite for relevant normative work, one of the more obvious ways philosophers can engage with scientists is by providing relevant and ultimately valuable normative recommendations (both in terms of how to "do" science and in a variety of other ways that range from how to aptly communicate scientific results to ways such results can be used in policymaking, for instance).

The resources needed to construct a descriptive model that can deal with these two challenges and prove useful in the context of a disciplinary shift towards practice-conscious work are present in recent (and less recent) work in philosophy of science, although they have not, as of now, been "put together". Kuhn's *Structure of Scientific Revolutions* (1970) outlines objections to traditional analyses of science which are similar to those that motivated the Practice Turn, although they are not discussed in a substantial way in the literature (with the notable exception of Rouse 2002). His construal of normal science as the solving of puzzles is particularly interesting, as it spurred the development of a family of theories of science centered on the related notion of problem-solving (somewhat indirectly in the case of Laudan and Hattiangadi). These theories offer the building blocks for a model of science that can accommodate the inherent pluralism that comes with the methodological project of the Practice Turn and, importantly, provide a way to deal with the challenges of specificity and normativity. I provide detailed reasons for taking the problem-solving approach in chapter 2. For now, I simply want to flag that my goal is to develop a descriptive model of scientific practice from the building blocks scattered through the literature on science as problem-solving.

My first step will be to propose a way to deal with the scope and normativity challenges outlined above. Chapter 2 introduces the notion of problem-solving as a basis for a descriptive framework of scientific practice. The idea is to show that scientific practice can be aptly described in terms of organized problem-solving activities. I begin with identifying six adequacy conditions that a descriptive model of science must meet if it is to function as an apt description of scientific practice, and I then review three models of science as problem-solving (Laudan 1978; Hattiangadi 1978, 1979; Nickles 1981) and argue that Nickles' is the only one that shows sufficient promise for my purposes. From there, I use Nickles' model to develop a general description of science as problem-solving.

Chapter 3 deals with the task of providing a description of how problem-solving happens in science. The description of science as problem-solving does not by itself provide an account of how it moves along. Moreover, while the problem-solving account deals with the scope problem, the issue of normativity remains unsolved. I argue that the best way to complete the model outlined in chapter 2 is to add the notion of heuristic, that is, the notion of an economical cognitive strategy that facilitates problem-solving, into the mix. After providing some arguments in support of this idea, I show how the notion of heuristic provides an account of how scientists solve problems, and how the resulting model of scientific practice not only meets the six conditions outlined in chapter 1, but also provides an avenue to resolve the potential issue with normativity: because heuristics are liable to systematic failures, it is possible, at least in principle, to identify when and how a heuristic goes wrong, thus providing a possible basis for normative recommendations. The chapter is rounded out by various illustrations of how heuristics are employed within scientific practice and by a preliminary statement of what I call the dynamic-iterative model of scientific practice. The model relies on the notions of problem-solving and heuristics as well as on a taxonomy of scientific problem and an account of how different types of problems relate to one another in a dynamic, iterative process that leads to the development of research agendas.

Chapter 4 introduces the dissertation's central case study: the case of the use of representational frameworks in neuroscience, in particular in the explanation of spatial learning and navigation. I first begin with an historical analysis of the role that the notion of representation played in the discovery of place cells, and then analyze the development of the place cell research agenda in terms of problem-solving and heuristics. I compare this analysis to that of Bechtel (2014) and show that the dynamic-iterative model allows for a richer understanding of the role representational frameworks play in neuroscience than Bechtel's model does. The application of the model to the case study is also meant as a demonstration that the model fulfills the six conditions outlined in chapter 2 and in particular the crucial requirement of being applicable in practice by philosophers of science.

Finally, chapter 5 begins with some objections to the model. Responding to these objections allows me to implement some conceptual revisions and clarifications to the dynamic-iterative model, in particular with respect to the taxonomy of scientific problems. These revisions made, I evaluate the model with respect to the six adequacy conditions and discuss its general strengths and weaknesses before moving on to comparing the dynamic-iterative model to other heuristics-based accounts of science, primarily Bechtel and Richardson's (2010). I spend the remainder of the chapter discussing the model's philosophical applications. In the concluding chapter, I discuss a potential practical application of the model, viz. its use as a framework for teaching philosophy of science to non-philosophers, which I take to be a promising first step in strengthening philosophy of science's potential impact outside of its disciplinary confines. This last task brings me all the way back to the initial motivations discussed in section 1.1. My aim is not *only* to develop a better descriptive framework for philosophers of science. Rather, my aim is to develop a framework that helps philosophers of science "export" their insights to other domains, and hopefully fulfill their potential as active collaborators not only within science, but within society at large.

Chapter 2

2 Problem-solving as a description of scientific practice

The aim of this chapter is twofold. First, I establish the conceptual foundations of the model I develop in the dissertation, which means I need to do two things: 1) explain how science can be aptly described as a problem-solving enterprise and 2) settle on a working definition of “problem’, by which I mean a definition that can be used to analyze the case study developed in chapter 4 and is flexible enough to be amended or otherwise revised following an analysis of its application.

Second, I want to argue that a problem-solving view of science is a natural fit for the development of a synthetic model of scientific practice that takes seriously the insights and methodological recommendations of the Practice Turn. I do this by showing that such a view provides a way to overcome the scope challenge discussed in the introduction, which is a consequence of the emphasis on empirical methods applied to case studies. Such methods are likely to be more descriptively accurate than the idealization-based methods criticized by proponents of the Practice Turn, but they may also struggle with making claims that go beyond these case studies. Simply put, construing scientific practice as organized activities, and describing these activities as problem-solving activities, makes it possible to provide a general description of scientific practice in terms of problem-solving.

The chapter is divided in four sections. I first provide some reasons as to why the general framework of problem-solving is the right starting point for understanding science, along with conditions of adequacy for the model (2.1). In sections 2.2 and 2.3, I discuss two models of science as problem-solving: Laudan’s (1978) and Hattiangadi’s (1978, 1979). These two sections have a shared structure: I first provide an overview of the model, identify its relevant contributions, and then identify its shortcomings. In both cases, the models are found wanting because they fail on (at least) the practical applicability and the descriptive accuracy conditions. This leads to the introduction of Nickles’ psychological view of problem-solving, which I argue avoids the problems of Laudan and Hattiangadi's models and provides a more apt description of scientific practice (2.4).

2.1 Preliminary considerations

The aim of this section is to lay out three reasons that justify the focus on problem-solving models. First, there is a rich body of work that shares a common origin with practice-oriented approaches, and indeed shares many concerns with critics of traditional philosophy of science, that begins with the idea that problem-solving plays an important role in the practice of science. One point of origin for this idea can be traced to Kuhn's *Structure of Scientific Revolutions* (1970), which outlines objections to the traditional analyses of science that resemble those that motivated the Practice Turn. While its importance for the movement is recognized, discussions of his work in relation to the Practice Turn are scant (Rouse 2002 being a notable exception). Kuhn's work did, however, inspire the development of all three problem-solving models I will evaluate in this chapter. Kuhn likened the practice of normal science to the solving of "puzzles", or "problems". Rouse explicates the analogy as follows: during normal science, certain well-defined "puzzles" constitute the bulk of scientific research and are assumed to have unique and accessible solutions, much like traditional puzzles do (e.g., crosswords puzzles). An apparent weakness of the analogy is that it paints "puzzles" as trivial and does not account for the intellectual significance we usually associate with the regular practice of science, thus making it disanalogous. The disanalogy, however, captures the way philosophy of science often misses the fact that the "big goals" of science are too removed from day-to-day activities to be motivationally significant (Rouse 2002, 104). Rouse argues that the solving of well-defined, bounded puzzles captures the way scientists pursue proximal goals that range from explaining a given phenomenon, to charting relevant regularities in phenomena (McMullin 1979), to securing grants or replicating experimental results and so on. This basic shift in the understanding of what it is philosophy of science needs to account for (i.e. the shift from providing accounts of science as a large-scale enterprise geared towards lofty goals to providing an account of the nitty-gritty, day-to-day practice of science construed as solving well-defined, small-scale puzzles) is coherent with the general methodological shift of the Practice Turn, and as such, looking at the approaches derived from this insight is likely to provide interesting avenues for developing a general descriptive scheme that fits with the methodological constraints of the Practice Turn.

Kuhn's views on the problem-solving aspect of normal science spurred the development of a family of theories of science centered on the notion of problem-solving (somewhat indirectly in the case of Laudan and Hattiangadi). I believe that problem-solving offers a promising basis for a model of science that can accommodate the inherent pluralism that comes with the methodological project of the Practice Turn. Thus, the first reason for focusing on those models is that since the literature on science as problem-solving developed around the same time as the Practice Turn and shares many of the goals and concerns of practice-oriented work, it is at the very least reasonable to think that the two may share as-of-yet unexamined similarities. I aim to show that synthesizing the methodological and disciplinary aims of practice-oriented approaches with those of problem-solving models is a promising avenue for developing a descriptively accurate model of science.

Second, the literature on problem-solving provides a characterization of practice that can be very useful in crafting accurate descriptions of science with which philosophers of science can work. Recall the definition of practice proposed by the Society for Philosophy of Science in Practice: "organized or regulated activities aimed at the achievement of certain goals" ("SPSP" 2017). Assuming this definition is at least in the ballpark of a relevant construal of practice in the context of science, then philosophy of science can be said to be interested in elucidating the kinds of activities involved in generating scientific knowledge and what considerations (internal or external to science) shape and determine these activities and the way they are carried out. One way this kind of methodology can be implemented is to develop a general characterization of the activities that make up scientific practice, one that is flexible enough to accommodate the range of activities that occur in various contexts but that also offers conceptual tools for describing and understanding these activities. A problem-solving model, which takes scientific practice to be made up of numerous instances of problem-solving, has the potential to fill that role. Each activity is construed as a particular problem-solving episode. Each episode occurs within a context, and from this contextual information it is possible to tease out the objectives towards which these activities are directed, and the "rules" (accepted methodologies, theoretical traditions, etc.) that constrain the problem-solving process. These episodes can be broken down further by identifying sub-problems that crop up in the context of the higher-level episode, making it possible to develop both fine-grained accounts of practice and more general accounts. This is how problem-solving models can overcome the challenge of the restricted scope of practice-oriented methodology: they can provide the conceptual tools for detailed, case-study-level accounts of scientific practice, while providing the conceptual tools to bridge those case-study-level analyses to general descriptions of scientific practice, by showing how the activities involved in solving the sub-problems relate to higher-level problems.

Third, the literature on problem-solving in science is complemented by a rich body of empirical work on human problem-solving. Given the importance of realistic descriptions of scientific activities for my project and, crucially, of accurate descriptions of scientists themselves, this knowledge is an important resource. So far, I have argued that construing scientific practice as made up of numerous problem-solving episodes is consistent with the aim of increasing the accuracy of our descriptions of science. These activities are performed by scientists, and scientists are human. The empirical literature on human problem-solving thus provides ample knowledge that is of direct relevance to the descriptive project of philosophy of science construed as the study of the problem-solving activities that make up scientific practice. In fact, philosophers of science have already tapped into psychology to develop accounts of explanation (Bechtel 2008) and discovery (Bechtel and Richardson 2010), explicitly developing these accounts as "naturalistic" accounts of science that rely on human psychology to define familiar concepts such as scientific rationality, giving central importance to the strategies employed by humans to solve problems (a significant chunk of which are heuristic in nature, see chapter 3) and how those strategies are used in scientific contexts. My contribution to this literature is to synthesize the literature on problem-solving with the insights of the Practice Turn and the recent work on heuristics in science. Thus, the third reason for the relevance of problem-solving models is that we already have a large body of work from which to draw accounts of *how* problems are solved.

These considerations are meant to set the stage for the remainder of the chapter, which focuses on providing support for the claim that problem-solving models can overcome the scope challenge while respecting the methodological constraints outlined in the introduction. A reminder of these constraints and their justification is required here, since I will be using them to assess the problem-solving models in the next sections.

Recall that descriptive accuracy is the overarching concern motivating the criticisms that "unites" the Practice Turn. This concern with descriptive accuracy translates to four shifts away from methods that rely on idealization of science and of scientists. First, theories are thought of as artifacts, tools that are evaluated not only on their logical structure or empirical adequacy, but also on their practical uses in a given context. The characterization of this shift needs to be amended so that the artifacts relevant to philosophical analyses of scientific practice are not limited to theories, since there are many areas of science in which there are no large-scale theories. Thus, this shift is perhaps best understood as the reinterpretation of theories as another tool that scientists can use in pursuing their goals, as opposed to the development of a theory being the only recognized goal of scientists (which does not preclude that scientists may well have the goal of developing a theory). From this shift, I derive the first adequacy condition for a descriptive model of scientific practice: it needs to be able to account for the ways theories and other tools are used by scientists and be sensitive to the pragmatic factors that may enter in theory-choice.

The second shift is a shift from *a priori* to empirical methods, which is to say that descriptions of science should be constructed through empirical investigation. This shift supports the condition that an appropriate model of practice needs to be flexible enough that it can be applied to a wide variety of cases and provide something akin to a framework, or a toolbox, for analyzing case studies, as opposed to a rigid model that would be imposed on a case study. Simply put, the shift to empirical methods comes with the desiderata that whatever descriptive scheme is used should be able to support empirical investigation, and thus should have some measure of flexibility so that it can be "fitted" to variable cases, as opposed to requiring that cases be fitted to *it*.

Third, there is a shift from idealized conceptions of scientists to more realistic, naturalistic accounts of scientists and of scientific rationality. From this I derive the condition that the model should give a psychologically realistic account of the activities that make up scientific practice. Putting this in terms of problem-solving, a model's account of how problem-solving happens should take into consideration known facts about human cognitive capacities and limitations. That is, the model should be a model *of* problem-solving, it should spell out what problem-solving *is*, and how it happens, within the limits of human cognition.

The fourth shift concerns the social dimension of science, which is given proper consideration, thus making historical and social context integral to the analysis of particular cases. The fourth constraint is that a good model needs to be able to accommodate relevant contextual elements of case studies, be they historical, social, cultural, etc.

To these four constraints, I want to add two further conditions. The first of these (fifth overall) should be thought of as a meta-condition that provides a general normative framework not only for the choice of appropriate solving models, but also for the development of any account of scientific practice. As such, the four conditions discussed above are in service of this condition which is that the problem-solving model should be useful *in practice*. That is, just as scientists are humans, with bounded cognitive capacities, so are philosophers. The model's usefulness must not depend solely on its theoretical applicability and descriptive power, but also, and crucially, on whether it can actually be implemented by philosophers of science. This means that the model must be useful in the sense of picking out relevant aspects of scientific practice as well as useful in the sense of not making unreasonable demands from its users in terms of time and cognitive resources. Moreover, my goal in developing this view of science as problem-solving is to construct a model of scientific practice that can be applied to cases old and current, one that can be put to work in academic and non-academic contexts. Thus, if a model is in theory very sound and very useful, it might still fail the in-practice applicability test. This is part of what motivates the use of a case study: I want to show that the model can be applied to an actual case. Note that while this condition has a different status than the other ones, I will treat it as one more requirement models of scientific practice should meet.

Second (sixth overall), the model should be able to support the development of an evaluative framework. That is, its descriptive resources should provide a way to evaluate choices made by scientists. That is, the model's definition of "problem" needs to be complemented by an account of how problems are solved, and that account of problem-solving needs to be one that makes it possible to evaluate historical choices made by scientists *and* that makes it possible to locate potential errors in contemporary cases. The model does not need to come with a full-blown normative framework (and in fact it probably should not), but it should support evaluative claims because the reason to develop such a model is to foster meaningful engagement between philosophers of science and scientists. Identifying potential errors can lead to interdisciplinary discussions about how to correct these errors given the aims of the researchers involved, the relevant disciplinary standards, and other contextual factors. In contrast, having a complete normative framework bypasses that discussion and can potentially hinder engagement between scientists and philosophers.

The six conditions are outlined in the table below. As a clarification, these conditions are derived from the four methodological shifts discussed above, as well as from the general aims of my project.

Table 1: Adequacy conditions on models of scientific practice

Account for the ways theories and other tools are used by scientists, and be sensitive to the pragmatic factors that may enter in theory-choice	Understanding the role theories play in scientific practice is necessary to ensure the descriptive accuracy of the model.
Support empirical investigation through accurate conceptualization and flexibility of application	The model and its key concepts should be flexible enough to capture scientific practice. This also applies to any conceptual tools developed, such as the

	notion of a problem, which can be defined in various way, some of which may not be accurate.
Naturalistic/realistic construal of human cognition	In the context of understanding how problem-solving functions, an accurate understanding of human cognitive capacities is necessary to ensure descriptive accuracy and relevance of potential evaluative and normative claims
Account for social nature of scientific practice	This condition is important because of the requirement for descriptive accuracy
In-practice applicability	This condition is important because if the proposed model is not realistically applicable by philosophers of science, then its practical value is nonexistent.
Support evaluative claims	Important for fostering interdisciplinary engagement

I believe these considerations justify my interest in problem-solving models insofar as together they suggest that problem-solving models are not only consistent with the general project of practice-conscious philosophy of science, but also come with considerable explanatory resources in the form of their connection to empirical work on human problem-solving. Moreover, I have outlined six desiderata that a model of science as problem-solving should meet if it is to be a promising basis for a synthetic account of scientific practice. In the next section, I review the first candidate model, developed by Laudan (1978) in the context of his theory of scientific progress.

2.2 Laudan

Laudan's *Progress and its Problems* (1978) outlines a well-known theory of scientific progress in which the notion of problem-solving plays a central role. My motivation for discussing it here is that while it is ultimately unsuitable for my purposes, it provides a useful point of contrast in that Laudan's theory is an example of a view that fails to meet the six criteria I outlined in the previous section.

Laudan's theory of scientific progress (1978) rests on the idea that "the single most general cognitive aim of science is problem-solving" (1978, 124). Importantly, this makes Laudan's view of science wholly non-epistemic, since he makes problem-solving not the means to attain knowledge, but the goal of science. This is a necessary feature of his account, given that his aim is to explicate the notion of scientific progress without falling into debates about epistemic values, preferring instead an epistemically neutral calculus based on the problem-solving effectiveness of competing theories. Appropriate theory choices are *progressive* insofar as they result in pursuing the theory most effective at solving problems. The idea is that explicating what constitutes progressive theory choice will serve to define scientific rationality. On this view, rationality is a *derivative* concept, as opposed to the foundational role it held in traditional approaches. Theory-choice is made on the basis of a comparative analysis of the problem-solving effectiveness of two competing theories (McMullin 1979, 627). As long as this choice results in the adoption of the most effective theory, the choice will be deemed progressive. The rationality of a given scientific community is derived from its ability to make progressive theory-choices, not the other way around[12]. Moreover, Laudan's theory maintains the focus on theory-choice characteristic of traditional approaches, although he seems to adopt a more nuanced view of it, asserting that scientists can hold a number of different attitudes towards theories beyond adoption and rejection (Laudan 1981; Hacking 1983, 15). Laudan's theory can be broken down into six theses.

---

[12] Since the assessment of a theory is always comparative, a theory is never abandoned if it is the only contender in the field, no matter how bad it is, which is a point criticized by McMullin (1979).

The first thesis is that “science is essentially a problem-solving activity” (Laudan 1978, 11), which on Laudan’s interpretation means that problem-solving is the most general *goal* of science. Laudan contends that while this is not a new idea, there has been no serious attempt to work out the consequences of this insight for theories of scientific *progress*. Laudan thinks that the notion of progress will be better understood if we begin with the idea that the goal of science is to solve problems (McMullin 1979, 624).

The second thesis attempts to provide a definition of problems. On Laudan’s view, there are two types of problems: empirical and conceptual. An *empirical* problem is anything about the natural world that seems to require an explanation[13]. Such problems are *about* the world, rather than about our theories of it. The precise nature of empirical problems will depend on our theories, as these will tell us what to expect from the natural world, and anything that does not fit these expectations will likely constitute an empirical problem. McMullin and Laudan refer to these as first-order questions (Laudan 1978, 48; McMullin 1979, 624). In contrast, *conceptual* problems are higher-order questions about the conceptual structures (e.g. theories) developed to answer first-order questions. They are problems *of* theories and can be internal or external. Internal conceptual problems result from challenges to the consistency or clarity of a theory, while external conceptual problems stem from tensions between a theory and some other theory or doctrine that proponents of the former are disposed to accept or at least recognize as well-founded (Laudan 1978, 49)[14]

Laudan’s third thesis is that there are two types of propositional networks that are commonly referred to as “scientific theories”, and that they must be distinguished from each other (Laudan 1978, 71). A *research tradition* is a general (and not as easily

[13] Note that this definition of problems as “something about the natural world that requires an explanation” will ultimately undermine Laudan’s theory, because he is unable to give a good reason why “problem-solutions” are not simply “explanations”, in which case all he’s done is add a new layer of analysis on top of the pre-existing approaches that took explanation of the natural world to be the aim of science (McMullin 1979, 636).

[14] Such tensions may have very different epistemic sources: other scientific theories, views about methodology and any component of a prevailing worldview (see McMullin 1979).

testable) set of doctrines and assumptions which at the most specifies a general ontology for nature as well as a general method for solving problems within a given domain. *Theories,* on the other hand, provide a more specific ontology as well as more specific and testable laws about nature (McMullin 1979, 626). Research traditions provide guidelines for the development of individual theories by providing a basic ontology and a set of appropriate methods for a given domain. The function of theories is to solve empirical problems in a given domain and what counts as a problem (and what counts as a solution to a problem) is defined by the research tradition. Moreover, the "cognitive loyalty" of scientists is to the tradition, not to individual theories, and the metaphysical and methodological tenets of the research tradition dictate what counts as an acceptable problem, avenue of research, etc. This thesis is particularly important for Laudan because it ties scientific progress to changes in research traditions as opposed to changes in theories. This way Laudan can propose new evaluative standards for research traditions, whereas proposing new evaluative criteria for theories would require showing that existing criteria are not adequate. The difference between a theory and a research tradition is thus one of generality and testability.

Finally, Laudan claims that theories and research traditions can only be assessed *comparatively*. This means that the only relevant questions that can be asked are about the *relative* problem-solving effectiveness of a theory (McMullin 1979, 628). As a consequence of this, Laudan holds that a theory is never abandoned unless there is an alternative to it[15]. Empirical anomalies (when a theory is in tension with a set of observations), when they happen, should not be counted against a theory unless the anomaly can be handled by a competing theory. If there is no alternative theory that can handle the anomaly, then it is a mere unsolved problem, which is not taken into account in the evaluation of a theory. Only solved empirical problems and anomalous problems, as well as conceptual problems, are taken into account in the evaluation process. Laudan further holds that anomalies can be eliminated through ad-hoc modifications of theories, which he sees as progressive as long as 1) there is no competing theory and 2) the

[15] Laudan thus rejects falsificationism entirely, as even in the face of an empirical anomaly, a theory will not be abandoned if it has no competitors.

problem-solving effectiveness of the original theory is not diminished as a result (Laudan 1978, 118). Laudan's view is that "ad-hocness" can be a cognitive virtue insofar as the modification to the theory is limited to enabling it to handle an anomaly and nothing more. Note that while such modifications, *if* they are in fact limited to handling the anomaly and have no other consequence for the theory, *are* strictly speaking empirically progressive, they may well introduce new conceptual problems in the theory, which in turn decrease the theory's overall problem-solving effectiveness (McMullin 1979, 629).

In summary, the basic unit of progress is the solved problem: progress is measured based on how many problems a theory solves, how effective it is at solving problems, and on an estimation of the theory's future problem-solving effectiveness. Progress is defined as an increase in problem-solving effectiveness. The problem-solving effectiveness of a theory is determined by assessing the number and importance of empirical problems the theory solves and deducting the number and importance of anomalies (that are handled by a competing theory) and conceptual problems the theory generates. Laudan wants to evaluate research traditions rather than theories, and the progressiveness of a research tradition is the combined problem-solving effectiveness of the theories that constitute the research tradition (Laudan 1978, 107). The rate of progress of a research tradition corresponds to changes in its adequacy over a given period. Now, the main advantage of this approach, for Laudan, is that it makes the process of estimating the progressiveness of research traditions a "simple" matter of calculus. In principle, one can determine all the relevant values by weighting and counting solved and unsolved problems and obtain a clear measurement of how progressive a research tradition is. Two further meta-theses are supposed to clarify *how* this method can be used by philosophers of science.

The first meta-thesis (thesis 5 overall) is that there is a body of "archetypal" comparative judgements that are universally accepted and can be used to test proposed models of scientific rationality. The test is supposed to evaluate the degree to which a theory of scientific rationality is "confirmed", or accurate. Here is an example: according to Laudan, an uncontroversial archetypal comparative judgement would be that it was rational to accept Newtonian physics over Aristotelian mechanics in the 1800's. A model of scientific rationality is applied to the case, and the philosophers figure out what recommendations the model would make. If the model applied is Laudan's, it would recommend the acceptance of Newtonian mechanics insofar as it would constitute a progressive (thus rational) theory-choice, presumably because Newtonian mechanics has an overall better problem-solving effectiveness than Aristotelian mechanics. On Laudan's view, the rational theory-choice is the progressive one, and in this case the model recommends the adoption of Newtonian mechanics because of its problem-solving effectiveness, which is comparatively better than that of Aristotelian mechanics. These recommendations are then tested against the historical record, yielding an evaluation of the model's accuracy[16]. The second meta-thesis (thesis 6 overall) is intended as a justification of the use of problem-solving efficiency in place of the more traditional criterion of how close a theory is to the truth. Laudan's argument for this is simply that we have no epistemically sound way of determining whether science is getting close to the truth (Laudan 1981).

2.2.1 Evaluation

A notable hiccup in Laudan's theory is that while he distinguishes between empirical and conceptual problems, there is scant detail on what a problem is. He adopts the view that a problem is anything that seems puzzling or peculiar considering our existing theories of nature (McMullin 1979, 635). In an attempt to clarify this, McMullin points out that a fact is puzzling only insofar as it calls for an explanation of some kind. Thus, it appears that on Laudan's view problems boil down to requirements for an explanation; the inability to explain something (e.g. the regularity with which heavy bodies fall toward the Earth) is what makes it puzzling in the first place. "Problem" is thus a derivative category: it is only because we seek to understand certain things that we label certain features of the natural world as "problematic". This poses two difficulties. First, defining problems as requirements for an explanation leaves out a large chunk of the work that scientists engage in regularly, such as the charting of empirical regularities or any other

[16] McMullin (1979) argues that this reverses the dependency relationship between progress and rationality that characterizes Laudan's view. Recall that he holds that rationality is a derivative concept.

activity that isn't explanatory (McMullin 1979, 636). The definition is simply too restrictive, especially given the kind of descriptive breadth required for my purposes.

The second difficulty stems from Laudan's distinction between solving problems and "explaining facts", a distinction that seems difficult to hold if a problem ends up being nothing over and above the requirement for an explanation. Laudan supports the distinction by arguing that problem-solutions differ from explanations in important respects, notably flexibility. His argument for this rests on using a definition of "explanation" that corresponds to the deductive-nomological (DN) model of explanation[17], which McMullin judges an unfair move insofar as Laudan picks a view of explanation that is easy to undermine and on which explanations are inflexible and rigid. First, Laudan argues that problem-solutions are approximate, but, argues McMullin, so are explanations. Further, Laudan points out that a problem-solution may cease to be a solution as new discoveries are made, but this is also something that happens with explanations. On McMullin's view, then, Laudan has not shown that problem-solving is relevantly different from providing explanations.

To be clear, the charge here is that Laudan's problem-solving model of progress is "merely" a layer of analysis that is added on top of explanation-centered analyses of science. If problem-solving boils down to explanation-giving, then the more progressive theory is the one that explains the facts the best. A likely response to this charge, and one I believe is convincing (to a point), is to point out that the problem-solving characterization is useful because it better lends itself to a non-epistemic, measure-and-count approach to the evaluation of theories. The idea is that the process of evaluating theories and theory-choices is made easier if the previously central notion of truth (i.e. the best theory is the truest one) is replaced by a measure of the progressiveness of the theory (problem-solving effectiveness). There is merit to this idea, as it is in principle possible to calculate the problem-solving effectiveness of a theory, a value that is more easily

[17] This is a classic view of explanation on which "explaining" roughly amounts to subsuming a phenomenon under a law through deductive reasoning, hence the deductive and nomological components (Woodward 2014).

measurable than the degree to which a theory is true[18]. The problem-solving analysis basically converts explanatory power into problem-solving effectiveness, which can be quantified and used in straightforward comparative evaluations of theories and research traditions. Trouble begins when considering what would need to happen for the model to be used in practice.

The applicability of Laudan's model relies on five assumptions. The first assumption is that problems can be unequivocally recognized as such. It is unclear how this would work given that different conceptual frameworks introduce different problems, some of which may not be problems for other frameworks. The second assumption is that one can determine whether a problem has been, in fact, solved, which differs from the weaker requirement of a problem being considered solved by a particular scientific community. The third assumption is that one can individuate problems sharply enough to count them. Fourth, it is assumed that one can reliably assign relative weights to solved empirical problems and to anomalous or conceptual problems and, further, that it is feasible to deduct the negative weight of the latter. Finally, it is assumed that one can determine the weight of each theory derived from a research tradition, including those that are incompatible with each other and combine the appraisals of these theories to figure out the problem-solving effectiveness of the research tradition. None of these assumptions are obviously true, but more importantly, each represents a task with considerable cognitive and temporal demands (McMullin, 1979, 637). Recall that one of the criteria for a suitable problem-solving model is that it needs to be *usable* by philosophers of science, and the foregoing casts serious doubt on the in-practice applicability of Laudan's model.

Another wrinkle stems from Laudan's focus on theories and research traditions. His model *requires* the possibility of clearly identifying theories and clear research traditions, because the entire point of the model is the quantitative comparison between research

---

[18] Moreover, one could convincingly argue that problem-solving effectiveness is an evaluative criterion that scientists themselves take seriously in their day-to-day activities, making the model more consistent with the actual practice of science.

traditions, which itself requires the quantitative assessment of their component theories. Such a model is workable (at least in theory) when dealing with areas of science that have large-scale theories and clearly identifiable research traditions, but this does not describe the entirety of science. Neuroscience, for instance, is a relatively new discipline, one in which large-scale theories do exist, but in a much looser way than, say, physics. Progress in neuroscience is technique-driven, with new imaging, recording and data analysis techniques generating not only empirical data, but also new conceptual frameworks such that identifying theories becomes even more difficult. For instance, the common assumption that 2-D images and 3-D objects are processed the same way by the brain can be construed as a form of theory in the context of experimental design. Technical innovation in the form of a plastic conveyor belt that could deliver real objects to test subjects inside the scanner lead to the discovery that the assumption was wrong, as the characteristic adaptation effect associated to 2-D images (increasingly lower neural response upon repeated presentation of the same stimulus) does not occur in the case of 3-D objects (Snow et al. 2011). This is an instance of technique-driven progress. Moreover, recall that for Laudan, a theory is a propositional network that specifies an ontology along with laws about nature. Such structures don’t show up in neuroscience, at least not in an easily identifiable way. Different laboratories develop different ontologies, models of cognitive and neurological processes (based on different ontological commitments, see, e.g. Sullivan 2009), and while there are undoubtedly similarities between some of these, these similarities are a far cry from the kind of propositional network Laudan takes theories to be. Similarly, neuroscience and other areas of biology do not necessarily have anything close to clear research traditions, although it is possible to identify a kind of theoretical lineage within which a research program might unfold, these lineages are a lot vaguer than research traditions, which would make their quantitative evaluation next to impossible[19].

[19] One can find theory-analogues in neuroscience. For instance, Goodale and Milner’s two-stream hypothesis (Goodale and Milner 1992; Goodale and Milner 2005) exhibits some features of a theory, but it does not provide a specific ontology nor specific and testable laws about nature, the way theories do on Laudan’s view. At this stage, the two-stream hypothesis is just that, a hypothesis, albeit an extensively studied and well-supported one.

In summary, Laudan's model is unsuitable for my purposes (see table 2 below). First, Laudan's model can handle the pragmatic factors involved in the use of theories by scientists, as evidenced by the claim that scientists may pursue (not accept) a theory riddled with anomalies if there is no alternative theory. The model fails to meet the second condition, however, since its definition of "problem" is far too restrictive given the descriptive scope I require. Its reliance on the notions of theory and research tradition compound this descriptive inadequacy. It also fails to meet the third condition: Laudan does not have much to say about how problems are solved, making the model somewhat irrelevant to the project of understanding scientific practice defined as problem-solving. While the model does seem to take into account the social aspect of scientific practice (fourth condition), it fails to meet the last two conditions: the model does not appear to be amenable to practical applications (as argued above) and its normative resources are unclear.

Table 2: Evaluation of Laudan's model

Account for the ways theories and other tools are used by scientists, and be sensitive to the pragmatic factors that may enter in theory-choice	Yes	At least to some degree, since there is recognition of the fact theories may be pursued for pragmatic reasons.
Support empirical investigation through accurate conceptualization and flexibility of application	No	Laudan's definition of "problem" is too restrictive and seems to reduce problem-solving to explanation-giving. Similarly, focus on theories and research traditions restrict the scope of the model.
Account of problem-solving consistent with naturalistic conception of human cognition	No	There is no substantive account of how problem-solving unfolds.

Account for social nature of scientific practice	Yes	Laudan allows for the influence of social factors both internal and external to science.
In-practice applicability	No	Applying the model is complex and requires dealing with potentially intractable tasks.
Support evaluative claims	No	Because of how difficult it is to apply the model, it is unclear what evaluative potential the model would have, although in theory it can at least evaluate historical cases.

The difficulties with Laudan's model can be traced to its construal of problem-solving as the *aim* of science. If it was not the case that problem-solving is supposed to be the aim of science, Laudan would not need to distinguish problem-solving from explanation. Instead, explanations could be one of many sorts of problem-solutions, which is something Nickles' model, discussed below, does allow for. At any rate, the main purpose of Laudan's problem-solving view seems to be to allow something like a calculus to evaluate theories, which enables one to dispense with normative notions like truth or rationality. It enables the specification of a methodology for understanding how theories are evaluated relative to each other that does *not* require one to specify a general evaluative notion other than the (in principle) easily tractable notion of problem-solving effectiveness. Unfortunately, in addition to the descriptive inadequacies outlined above, "[instead] of a logic of confirmation […] we have a new and (it would seem) much more complex formalism to cope with" (McMullin, 1979, 637). I now turn to the second candidate model.

2.3 Hattiangadi

Laudan's theory of scientific progress fails to provide a workable construal of science as problem-solving for my purposes because its focus on problem-solving as the *goal* of science, in addition to its restrictive definition of what a problem is result in a model that is not only descriptively inadequate, but also difficult to apply in practice. While Hattiangadi's model is unsatisfactory in some respects, it does come with a few important insights which are crucial to the development of the model I discuss in 2.4. I begin with an overview of Hattiangadi's problem-based epistemology of science, with an emphasis on the components of his definition of "problem", and concludes with a critical evaluation of the theory.

Hattiangadi (1978, 1979) attempts to ground an epistemology of science on the notions of problem and problem-solving. His focus is on theory-evaluation and theory-choice, in accordance with what Giunti characterizes as the standard postulate of epistemologies of science:

> (TC) (under ideal conditions) scientists choose theories by following methodological rules, which are based on evaluation-concepts. (Giunti 1988, 422).

Hattiangadi aims to provide a way to specify evaluation-concepts by grounding them in the notions of scientific problems and problem-solving. His theory of problems is intended as a response to two difficulties that arise from views such as Popper's, on which we can never gain full, secure knowledge but nevertheless continue to pursue it (Giunti 1988). The first difficulty with this is that there seems to be no reason to pursue knowledge and, second, if we never truly have knowledge, how can we recognize that certain hypotheses are worth considering, or at least testing? Hattiangadi's claim is that these are easily solved by adopting a problem-based view of science: we seek knowledge because we have to in order to solve problems that arise within our beliefs. Moreover, problems themselves impose conditions on what counts as a solution on Hattiangadi's view, thus resolving the second difficulty (Hattiangadi 1978, 351; Giunti 1988, 423). Hattiangadi also insists on the importance of the historical structure of problems, arguing that the historical context in which a problem is identified and solved should be taken into account in the evaluation of theories (there is no decontextualized evaluative criterion). Finally, while Hattiangadi provides a clear analysis of the logical structure of problems, this structure ends up being too restrictive. This, in addition to issues of internal consistency (which I do not discuss in detail here, but see Giunti 1988, sec. 3.1) ultimately disqualifies Hattiangadi's theory. This being said, Hattiangadi's emphasis of the historical structure of problems highlights a crucial component of the view I develop in the next section.

First, Hattiangadi distinguishes two categories of problems. Intellectual problems share a logical structure, viz. they are all *contradictions* (more on this later), but some problems, e.g. scientific problems, have, in addition, *depth*. This is because scientific problems crop up in theories embedded in intellectual traditions, which gives them a historical structure that needs to be taken into account when analyzing them (Giunti 1988, 423). Intellectual traditions are lineages of theories and problems that originate from intellectual debates. If traditions stem from the same debate, they are *competing* traditions. An intellectual debate originates from a problem arising in a given theory and each possible solution to that problem is the starting point of a distinct tradition (Giunti 1988, 423). This is represented in Figure 1. Hattiangadi provides a formal illustration of the origin of an intellectual debate:

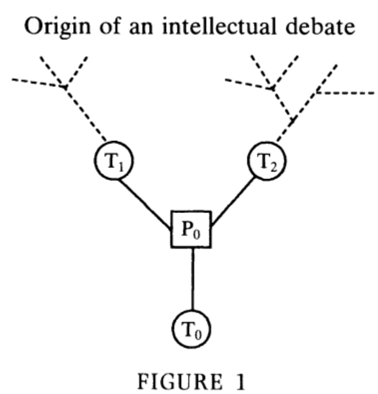


Figure 1: Origins of intellectual debate (Giunti 1988, 424)

In this figure, a theory T_0 gives rise to a problem P_0. Theories T_1 and T_2 propose competing solutions to P_0, and generate further problems, which are solved by further theories, and so on. The same debate at a later stage is represented in Figure 2.

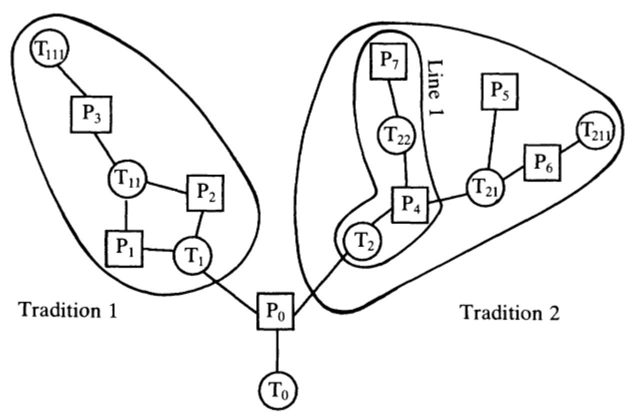


FIGURE 2

Figure 2: Debate from fig. 1 at a later stage (Giunti 1988, 424)

This figure illustrates how theories give rise to intellectual traditions, which include new problems that arise from theories as well as subsequent theories. Within traditions are found *lines of thought*, which can be open or closed depending on whether all problems within a line are solved. Closed lines are preferred over open lines. There is a complication here: a problem in a line may be solved but still *open* in the case where two problems stem from the same theory but only one of them is solved. The solved problem will remain open, since there is no theory that solves both it and the unsolved problem (Guinti 1988, 425-6). Hattiangadi remains ambiguous about this distinction, but it seems relevant to mention it here, as it is a consequence of Hattiangadi's claim that a problem always includes in its structure every problem that came before it in the same line of thought, which he dubs the property of continuity of scientific problems (Hattiangadi 1979, 55; Giunti 1988, 426). Moreover, one can easily see that the same problem never appears twice in a line (or for that matter, in the entire graphical representation), since problems are taken to be non-recurring on this view.

On Giunti's view, however, both properties should be abandoned because they simply are not properties of actual scientific problems (Giunti 1988, 429). Per the non-recurrence property, if a problem P is solved by a theory in a given line of thought, P can never occur in that line again as a problem for a subsequent theory. Giunti asserts that this is too strong a claim, as it rules out cases such as the problem posed by mutations for Aristotle's biology. Since cases of mutations are cases where an organism diverges from the essential characters of its species, Aristotelian biology faces a problem in explaining such occurrences. According to Giunti, the fact mutation is also a problem for Darwin's theory – since his theory has no clear account of the mechanism through which mutations occur – suggests that scientific problems do recur. I believe this criticism stems from a misunderstanding of Hattiangadi's claim. Hattiangadi may simply be emphasizing the importance of the historical context in which a problem occurs. On this view, the problems faced by Aristotle and Darwin differ in their contextual features, related though they may be. Note that this feature of problems – that they are nontrivially *defined* in part by their historical context, is a key feature in the view of problems I adopt later in the chapter.

Per the conservation property, if a given problem includes every problem that came before it in the same line, then any theory which solves the first should solve the others as well. Giunti argues that this is an odd claim to make because while one can trace a line between, say, Parmenides' problem of motion and Newtonian mechanics, the latter hardly seems to be solving the former (Giunti 1988, 430). Once again, this criticism downplays the historical aspect of Hattiangadi's view: if two problems are in the same line of thought, then the later problem will include previous problems as part of its historical structure. That is, the problem includes constraints that applied to the previous problems, and which would be absent were it not for these older problems. In fact, one can argue that insofar as problems are always encountered within the context of science *in general*, there is a sense in which every scientific problem figuratively carries DNA from all previous scientific problems, however minutely (Nickles 1981 makes a similar point). Neither of these objections seems insurmountable once the properties of conservation and non-recurrence are adequately understood in terms of their historical dimension. Giunti does, however, highlight a further issue with Hattiangadi's model: it is difficult to trace a clear historical line between temporally distant theories or problems. This constitutes a serious obstacle to the in-practice applicability of Hattiangadi's model. If such a concern was part of the reasons to reject Laudan's model, it should carry just as much weight here.

Hattiangadi's analysis of the historical structure of problems is a crucial component of the model I develop in later chapters. His analysis of the logical structure of problems, however, faces significant issues. On his view, all intellectual problems are logical inconsistencies, with scientific problems being distinct only because their historical structure gives them a depth that other problems lack. The issue here is that many problems encountered in science simply are not logical inconsistencies. Gaps in knowledge certainly qualify as problems, at least insofar as they constitute a large portion of scientific work, and they are not logical inconsistencies. That we know the mammalian brain enables spatial navigation but not *how* is a problem, but it isn't a problem because the lack of an account of how the brain achieves this entails a logical inconsistency. Similarly, measurement problems, such as when we know a given object has a physical magnitude of some sort but not what the value of that magnitude is, seem to be the kind of thing that should be included in the category of scientific problems, but they also are not logical inconsistencies (Giunti 1988, 430)[20]. Considering this, Hattiangadi's analysis of scientific problems lacks descriptive accuracy.

---

[20] Hattiangadi's technical definition of the logical structure of problems leads to the conclusion that there can be no solution to a problem which would in turn lead to another problem, which is a difficult position to hold since the fact solving a problem in a given way can lead to new unforeseen problems is a fairly uncontroversial claim. I omit the details of this point here as it requires diving into the formalism of Hattiangadi's theory, but a complete discussion can be found in Giunti 1988, 431-2.

Finally, recall that Hattiangadi’s objective remains to provide an account of theory-choice, the difference being that he wants to ground the evaluation-concepts used in theory-choice in the notion of problem. However, as was discussed in chapter 1, this focus on theory-choice is too narrow, as the bulk of scientific activity is concerned with tasks other than theory-choice. Moreover, his theory-centered analysis faces the same descriptive difficulty as Laudan’s: many areas of science (e.g. neuroscience) do not have large-scale, clearly defined theories. This being said, Hattiangadi’s emphasis of the historical structure of problems is a crucial contribution to the development of a workable problem-solving model of science, as will become clear in the next section.

Table 3: Evaluation of Hattiangadi's model

Account for the ways theories and other tools are used by scientists, and be sensitive to the pragmatic factors that may enter in theory-choice	No	Theories are still construed as the end products of scientific work, as opposed to tools that are used in practice.
Support empirical investigation through accurate conceptualization and flexibility of application	No	Definition of problems as logical inconsistencies excludes a large part of scientific practice.
Account of problem-solving consistent with naturalistic conception of human cognition	No	There is no substantive account of how problem-solving unfolds.
Account for social nature of scientific practice	No	Scientific problems being defined as logical inconsistencies with added depth conferred theoretical traditions limits the scope of the model to “pure” science with no room for social factors.

In-practice applicability	No	Applying the model is complex and requires dealing with potentially intractable tasks.
Support evaluative claims	No	No discussion of evaluative components.

2.4 The constraint-inclusion model: defining problems for practice

At this point, I have rejected two analyses of science as problem-solving. In both cases, the rejection was predominantly motivated by the fact the model fails to meet my practical applicability and descriptive adequacy conditions. This idea is central to the view of problem-solving in science I discuss in this section. I begin with an overview of Nickles' constraint-inclusion model of scientific problems. I then discuss two potential issues with the model and argue that neither pose serious difficulties. The next step is an evaluation of Nickles' model on the six conditions outlined in 2.1, followed by a demonstration of how the model can deal with the scope challenge.

Nickles' view of scientific practice (1981, 1974, 1978) presents a number of important differences from those discussed above. First and foremost, Nickles takes problem-solving to be an apt *general description* of the *process* of science, recognizing that the crucial step in developing a descriptive account of science so understood is to have a clear definition of what a problem is (1981). Constructing a satisfactory definition of problem is a necessary step towards constructing a satisfactory account of how scientists solve problems. Recall that Laudan's focus was on constructing an evaluative framework on the basis of the idea that science's aim is to solve problems, and Hattiangadi emphasized the development of new evaluation concepts based on a notion of problem-solving, and both of their definitions were impoverished as a result of putting the evaluative work first. Nickles emphasizes description as a stepping stone towards evaluative work. Description, on Nickles' view, is a serious task, not a mere formality. One of Nickles' main criticisms of approaches such as Laudan's and Hattiangadi's is that neither provide a clear account of problem-solving as a *process* (Nickles 1981). Although he also does not provide an account of such a process (he does gesture towards one), he insists on the crucial importance of starting with a clear understanding of what a problem is. It is, in fact, easy to draw a line between Nickle's constraint-inclusion model and the Practice Turn, especially with respect to concerns with descriptive accuracy. Recall from the previous chapter that one of the distinctive shifts of the Practice Turn was a move from idealized representations of scientists to a more accurate construal of their capabilities, behavior and motivations coherent with results from empirical psychology along with an increased importance of historical and social analysis. The importance of Nickles' contribution to the literature on problem-solving in science cannot be overstated. His work had a profound influence on Wimsatt's work (Wimsatt 2007a, 2006b, 2006a), which in turn was a crucial inspiration for contemporary problem-solving and heuristics approaches to science, such as Bechtel and Richardson's (2010). This influence began with his careful work in defining problems, which I turn to now.

Nickles asserts that a problem consists in "*all* the conditions or *constraints* on the solution plus the demand that the solution (an object satisfying the constraints) be found" (Nickles 1981, 109, emphasis in original). On this view, then, a problem is individuated and defined in terms of its admissible solutions. Nickles insists that constraints are the key elements that provide the definition of a problem, because the constraints specify (in a descriptive sense) the solution to a problem. This is an important feature of Nickles' account. It also leaves out an important aspect, viz. that of the constraints on the *search* for a solution, but this omission is easily remedied. Moreover, one could simply say that constraints on a solution include constraints on the means one can use to arrive at this solution and be only slightly too loose with the terminology.

Some precision is required here. First, Nickles points out that problems arise when certain goals or demands are formulated. They are not solved for their own sake; a problem does not exist in the abstract without a goal to achieve. Contrast this with Laudan's claim that problem-solving itself is the aim of science, a claim he had to make to ensure the relevance of his model for theory evaluation, which relied on the identification of epistemically neutral features of theories that could be used in comparative judgements. On Nickles' view, problems come up in the process of pursuing certain goals, and attaining those goals requires solving problems. Applied to the project of providing a description of scientific practice, this view provides a clear way to explain the motivations of scientists in solving problems: they stand in the way of the goals they pursue. This goal-dependent definition of problems makes it possible (in fact, requires) to identify the goals pursued by scientists, and then the problems that come up in the pursuit of these goals. It grounds a flexible descriptive scheme, which is particularly useful for the kind of descriptive project I am pursuing here.

Next, any (relevant) problem must include at least one constraint, lest it be entirely impossible to solve (without at least one constraint, a problem has an infinite number of admissible solutions and inquiry becomes impossible, at least in practice). Unconstrained problems are possible, but they represent an exemplar of problems that need to be constrained if a solution is to be found given limited temporal and computational resources. Nickles is clear that the general set of all possible constraints is quite large, but that for practical reasons, only constraints that are specific to a problem are usually made explicit in scholarly communication or everyday problem-solving (I refer to such constraints as "special" constraints). For instance, any problem will have as a constraint that the Earth is (roughly) round but, in most cases, such a constraint won't need to be made explicit. On the other hand, the constraint that internal cognitive states are not admissible as components of explanations of behavior is specific to explanatory problems tackled from a strict behaviorist perspective, and this last constraint also constrains the solution to the problem of choosing the most financially advantageous insurance policy, but it's not a *special* constraint of that problem. Moreover, such explanatory problems *also* include the "Earth-is-round" constraint, but not as any kind of special constraint. Thus, Nickles' model can handle any individuation scheme so long as it rests either on constraints or goals (Nickles 1981, 109). On the face of it (and by Nickles' own admission, see 1981, 109), this is similar to the view put forward by erotetic logicians such as Belnap and Steel that questions include their possible answers such that the logic of question-answering boils down to selecting a possible answer from the set offered by the question itself (Belnap and Steel 1976, 17). For instance, questions like "what is the color of Kramer's house?" provide a set of possible answers through simple discursive logic (the answer to the question will be, in normal circumstances, a color name).

This definition of problems has the advantage that it leaves nothing out of the category of problems that one might want to include. Laudan and Hattiangadi's definitions of problems were either unclear or excluded things like measurement problems or knowledge gaps from the category of problems. Nickles' definition can easily accommodate these instances, in addition to straightforward explanatory problems and problems internal to theories (e.g. internal consistency problems). It can also accommodate conceptual problems stemming from tensions between a particular theory or hypothesis and factors external to science. Importantly, the model can handle practical problems such as those that occur in experimental contexts. An example will be useful here. Imagine you're a neuroscientist engaged working on the visual perception of objects. Your overarching goal in this case is to explain how human brains process visual information about 3-D objects. In the process of attaining that goal, you'll face the problem of explaining what the neural correlates of perceiving an object are (explanatory problem). To solve that problem, you need to record activity in the brain while a subject perceives an object but, fMRI scanners aren't object-friendly (practical problem). You then decide to show subjects 2-D images, assuming that perceiving 3-D objects has similar enough neural correlates to perceiving 2-D images. This in turn, introduces new problems: there will be a need to ascertain the validity of that assumption, which itself comes with practical problems, and so on[21]. Each of the problems listed here can be grouped into categories (explanatory, practical, etc.), but because of the constraint-inclusion model's way of individuating problems, each problem within a given category can be clearly differentiated from other problems of the same category.

On the other hand, the definition might be including too much: anything can be thought of as a problem provided it includes the specification of a goal state and at least one constraint (implicitly or explicitly). This extensional vagueness is a double-edged sword

[21] This example is inspired by the work of Snow et al., who tested the assumption that perceiving 2-D images has the same, or similar enough, neural correlates as perceiving 3-D objects (Snow et al. 2011).

of sorts, but one I believe is ultimately more beneficial than harmful. In particular, the conception of problems on offer here can accommodate evolving case studies, identifying new problems and constraints as they occur, in addition to providing conceptual tools for detailed historical case study analysis. At the same time, the model is fairly general, and can be applied at varying levels of generality.

A possible objection to the view as developed so far is that it runs into a serious epistemic issue: puzzlement does not arise because we don't know which problem-solution to choose from a set of possible solutions, but precisely because we can think of no satisfactory answer. In addition, it seems difficult to distinguish between problems that have the same class of solutions. Nickle's solution is to show that we can know what counts as a solution to a problem without having the solution itself, in virtue of the fact that for us to even have a problem in the first place, we need to have *some* constraints on a solution. These constraints (and those discovered through inquiry) circumscribe the range of acceptable solutions, but they do not, and this is important, give the solution itself. In effect, Nickles' response to this objection is that there only seems to be a difficulty because the model can be misunderstood as holding that problems are defined by their solutions. In fact, problems are defined in terms of the things that determine the range of acceptable solutions, which are the constraints. Thus, inquiry is possible because we have guidance in the form of constraints on solutions, such that we can recognize what we're looking for when we find it (Nickles' answer to Meno's paradox). It follows, also, that problems are individuated based on their constraints: "two problems differ if and only if (and insofar as) their constraints differ" (Nickles 1981, 111), with the provision that distinct constraint sets can determine the same (or at least overlapping) ranges of admissible solutions.

On Nickles' view, then, a problem is a demand that a certain goal state be achieved coupled with constraints on the way this goal state is to be achieved, understood as conditions of adequacy on solutions and on the search for solutions (Nickles 1981, 111). The constraints may not all be known, and additional ones can be discovered through inquiry. Moreover, problems have an objective existence and can remain undetected, as they exist within a body of belief, assumptions and practices and goals regardless of whether they are known or not. For instance, problems which were never addressed by a theory can be discovered by newer theories. Constraints can also be *tacit*, such that they can influence the search for a solution. Some constraints are more fundamental than others, or better established than others (e.g. methodological standards in a given discipline).

Before moving on to the avenues this approach opens, it seems appropriate to bring up some important features of the constraint-inclusion model. First, while it does not call into question the epistemic nature of the aims of science (which Laudan does), it can remain entirely agnostic about what, precisely, those aims are. This is because Nickles' model does not require science to have any specific goals as long as it *has* goals at all. They could be epistemic, technological, or even particular to scientific communities (such as laboratories) but either way the only thing the model requires is that scientists pursue some sort of goal. On this view, problem-solving is a *description* of the activities scientists engage in while pursuing their goals. This is coherent with the idea of scientific practice as a set of organized activities as defined by the SPSP's mission statement.

A second distinctive feature of Nickles' model is that the definition of problems it is built on avoids the scope issues of definitions such as Hattiangadi's, for whom problems are only logical inconsistencies, and Laudan's insistence on excluding questions and explanations from the category of problems. Finally, another important feature of Nickles' model is the *inclusion thesis.* Nickles borrows this idea from Hattiangadi (1978) in order to reject claims made by Popper (1972) and Lugg (1978) that problems can be separated from their backgrounds, or settings (Nickles 1981, 94-5). In a nutshell, the inclusion thesis holds that a problem's historical setting, which includes the common conceptual frameworks, the state of the discipline, etc., is part of what determines the constraints that characterize it. Part of Nickles' argument for this is that the inclusion thesis enables one to make sense of evidence that conceptual constraints are constitutive of problems. For instance, he notes that there are many different types of problems, some of which are purely conceptual and do not (directly) involve empirical data (Nickles 1981, 88). I take the inclusion thesis to highlight a crucial feature of scientific problems. These features of Nickles' account make it a strong candidate for developing a general model of science in terms of problem-solving that is compatible with the insights of the Practice Turn, as it naturally draws attention to the importance of describing problem-solving behavior in science to account for how science functions. This focus leads scholars following in Nickles' footsteps to take into account the psychological factors involved in problem-solving, and develop a rich literature around the notions of problem-solving and heuristics.

Nickles' model appears to possess considerable descriptive power, but this descriptive power can be enhanced if the model is extended and developed further so that it meets the six adequacy conditions I laid out in 2.1. Note that it already handles some of the conditions quite well as-is (see table 4 below). First, the model has no trouble handling the requirement that theories and other artifacts be analyzable in terms of their practical use– as tools and not ends in themselves. Of course, the development of a new theory, a new predictive model, or a new recording technique can be analyzed as a goal, and the problems that occur in the pursuit of that goal can be clearly and usefully described using the constraint-inclusion model. However, the set of relevant problems is not limited to the development of theories or tools, and the model can characterize a theory through its *use* as a problem-solving tool. Second, the model is flexible enough that it can apply to cases in different areas of scientific practice. The constraint-inclusion model can both characterize large-scale scientific enterprises and small-scale work in a single laboratory. Most important is the fact that its definition of "problem" is flexible enough to capture the range of activities that make up scientific practice, something the previous models failed to accomplish. Third, its focus on the *process* of problem-solving makes it compatible with a range of accounts of how problems are solved. While it does not, by itself, meet the descriptive accuracy condition with respect to the cognitive capacities and limitations of scientists, it can be supplemented with an account of how problem-solving occurs that *would* meet that condition. Fourth, problem-solving can easily be thought of as a social enterprise, and the constraint-inclusion model is built with the importance of cultural, social and historical factors in mind, especially when it comes to the two key notions of constraints and goals – both of which are likely to be at least partially context-dependent.

Table 4: Evaluation of Nickles' model

Account for the ways theories and other tools are used by scientists, and be sensitive to the pragmatic factors that may enter in theory-choice	Yes	The model can describe the pragmatic role of theories in the process of problem-solving, while also being able to analyze theory development.
Support empirical investigation through accurate conceptualization and flexibility of application	Yes	The model is flexible enough to capture the range of activities that make up scientific practice, and can characterize different levels of practice.
Account of problem-solving consistent with naturalistic conception of human cognition	No, but potentially	The model does not meet the condition by itself, but is compatible with many accounts that would. Moreover, Nickles gestures towards a possible complementary account based on the notion of heuristic.
Account for social nature of scientific practice	Yes	The key notions of constraints and goals are highly sensitive to social and historical factors.
In-practice applicability	No	No demonstration of applicability in Nickles' discussion.
Support evaluative claims	No	No demonstration of evaluative potential.

Showing that the model meets the last two conditions, viz. that it is applicable in practice and provides at least the possibility of a normative account, requires work that is beyond the scope of this chapter. Specifically, the next chapter will focus on developing an account of the process of problem-solving, and I will argue that this account does have the resources to support normative frameworks. The fourth chapter is an application of the resulting model to a case study, which itself, I will argue, constitutes a demonstration of the model's in-practice applicability. Thus, Nickles' constraint-inclusion model warrants further consideration because it holds promise for developing a descriptively accurate account.

2.5 Conclusion

I began this chapter by providing three preliminary arguments to justify my interest in problem-solving as the starting point for a descriptive framework that could accommodate the insights of the Practice Turn while avoiding the scope challenge that comes with the methodological recommendations associated with those very insights. I then specified six conditions that a problem-solving descriptive framework should meet to qualify as a promising candidate for my project.

The first model I discussed (Laudan's) fails to meet several of these conditions, most notably the conditions of descriptive accuracy and in-practice applicability, although it does make an important contribution in the form of its recognition that scientists can hold a range of attitudes toward theories beyond accepting or rejecting them, which is a step in the direction of a realistic portrayal of scientists as agents engaged in the activities that make up scientific practice. Hattiangadi's model fails on much the same grounds as Laudan's, but his analysis of the historical structure of problems provided crucial insight that directly influenced Nickles' constraint-inclusion model. I then provided a description of the constraint-inclusion model.

The constraint inclusion model provides a way to overcome the scope challenge. Because it provides a flexible enough descriptive scheme, it can be applied to large-scale analyses of scientific practice (say, at the level of theory development over a long period of time) or to more fine-grained analyses (for instance, how a particular study solves problems that come up during experimental design). By allowing for the description of practice in terms of problem-solving regardless of the level of analysis, the constraint-inclusion model makes it possible to treat apparently disparate studies of scientific practice as only different in terms of the grain of the analysis, and to understand how small-scale studies may related to larger-scale analyses of practice. This flexibility is largely a result of Nickles' definition of "problem", which is flexible enough that it can both capture, say, the practical problems involved in developing an experimental protocol and the conceptual problems that drive the development of theories. Of course, this flexibility comes with a certain extensional vagueness with respect to the category of "problems", as well as the need to pay close attention to the numerous sources of constraints – be they contextual factors linked to the historical and social setting in which the problem was occurrent, or more explicit constraints willfully imposed by the scientific community.

At the very least, I believe I have shown that the constraint-inclusion model is a promising descriptive scheme that characterizes scientific practice accurately and is not tied to a specific level of analysis, thus making it a strong candidate for a general description of scientific practice. The next chapter tackles the issue of elaborating an account of how problems are, in fact, solved by scientists, one that conforms to the descriptive accuracy condition with respect to the psychological traits and limitations that scientists work with. Nickles himself suggested that whatever problem-solving ends up being, it is likely that the notion of heuristic would play a role, and I intend to show he was correct in this.

Chapter 3

3 Heuristics, description and prescription

My aim for this chapter is to argue that adding heuristics as a key component of the account of how the process of problem-solving occurs in science completes the picture in a way that makes the constraint-inclusion model very well-positioned to serve as a general descriptive framework for philosophy of science, at least in theory. Moreover, I aim to show that this account would allow the model to overcome the normativity challenge associated with the methodological recommendations of practice-oriented approaches. Recall that their rejection of traditional accounts based on abstract principles of logic and reasoning meant they were cut off from the "free" normative account that comes with that kind of view and thus had to secure normativity from somewhere else. I believe that a heuristics-based approach can provide the normative resources needed to ensure that philosophers of science can make actionable recommendations for practitioners and users of science.

The chapter is divided in three sections. In section 3.1, I provide some justification for my focus on heuristics as a core mechanic of scientific practice. I argue that the notion is particularly relevant to the project of giving an accurate description of the activities that make up scientific practice, in large part because of its entrenchment in the psychological literature following the work of Kahneman, Slovic and Tversky (1982) and the subsequent development of accounts of human cognition that focus on the "fast and frugal" heuristics that enable economical decision-making (e.g. Gigerenzer et al. 1999). Section 3.2 focuses on explicating the notion and translating it into the context of scientific practice, a task made easier by the fact there already exists a sizeable literature on heuristics in science (e.g. Bechtel and Richardson 2010; Wimsatt 2007). This literature is, in parts, motivated by the same kinds of concerns that motivated the Practice Turn, but their scope is often more restricted than mine in the sense that the combination of the constraint-inclusion model as a description of practice and of heuristics as a core component in the *process* of scientific practice provides a much more flexible, and general, descriptive framework. Finally, section 3.3 provides a first formulation of my model of scientific practice. This section will also specify how the model meets five of the six conditions laid out in chapter 1, and map out how the task of testing whether it meets the sixth condition (in-practice applicability) will be accomplished in the remainder of the dissertation.

3.1 Why heuristics matter

One of the arguments given in support of the relevance of problem-solving models in chapter 2 was that there is an extensive empirical literature on human problem-solving. In his 1981 paper, Nickles explicitly acknowledges the influence of Simon (1977) on the development of his account of problems, emphasizing the importance of *heuristics* for the task of understanding how problems are solved, citing additional work by Newell, Shaw, and Simon (1962) in which they explicitly draw a link between their conception of problem (which is very close to Nickles') and the role of heuristics in problem-solving (Nickles 1981, n. 17). In a nutshell, "heuristic" is commonly used as a label for certain procedures used to guide inferences or judgements in problem-solving and decision-making (Chow 2015, 978). What's interesting about heuristics is that they function as *shortcuts*, insofar as such procedures may produce merely "good enough" results as opposed to optimal results. They are also more economical than thorough methods, requiring less time and energy. On the flipside, they are likely to be biased, resulting in systematic failures in certain contexts. Importantly, such biases can be identified, which provides an avenue for the development of normative accounts based on this empirically informed description of problem-solving as a process. I later review some of the evidence suggesting that humans commonly employ heuristics, but simply put, heuristics allow human agents to solve problems and make decisions without spending considerable time and energy each time they need to solve a problem.

My goal is to provide an accurate description of scientific *practice*. This means describing not only the end products of science (e.g. theories) but also the steps required to arrive at those end products. This, in a sense, replaces what was once taken to be the primary goal of philosophy of science: instead of being tasked with elucidating the logic of justification for theories, the philosopher of science is tasked with understanding how scientific practice works. When the focus is put on scientific practice, the first step is to provide a definition of "practice". I provided this general description in Chapter 2, understanding scientific practice as consisting of many instances of problem-solving activity. If scientific practice is "scientific problem-solving", then the job of philosophers of science is to explain how problems are solved in science. One important methodological shift discussed above is the shift to a realistic construal of scientists themselves, and this is where heuristics come in.

A further motivation for focusing on heuristics is drawn from the fields of epistemology and philosophy of mind. With interest in heuristics building up in psychological circles, naturalistic accounts of human epistemic and reasoning capacities have taken up the idea that the human mind being a product of evolution, it is likely that human cognitive capacities are, to some extent, shaped and limited by the parameters of said evolution. Note that I do not wish to include work that falls under the "Evolutionary Psychology" umbrella here, as such work is characterized by a specific set of assumptions that are not necessarily shared by every scholar working from an evolutionary perspective. Such work either puts forward very strong theses about cognitive architecture which go far beyond the scope of simply tackling epistemological questions from an evolutionary standpoint (see for instance Tooby, Cosmides, and Barrett 2005) or are of a more programmatic nature (Barkow, Cosmides, and Tooby 1992)[22]. The sort of work I have in mind here takes the (defensible) assumption that because human brains have evolved over time, it is likely that whatever cognitive capacities they have are supported by structures layered over pre-existing structures. A further assumption is that optimization is not something that evolutionary processes are particularly concerned with, preferring economical solutions that mostly work, and that the end result is that the human mind is a collection of "good enough" structures that support "good enough" cognitive capacities (see Marcus 2009). In Wimsatt's terms, on this view nature is thought of as a kind of backwoods mechanic "patching up" as it goes, producing systems that get the job done (Wimsatt 2007, 9–10). This view has inspired work on reasoning (Clarke 2004) and on perception (Viger 2006) to take into account the likely possibility that humans are not

[22] This work has been harshly criticized from several directions (Lloyd 1999; Lloyd and Feldman 2002; Richardson 2010).

"designed" for optimal functioning, but rather for "good enough" functioning in the context of enabling survival.

Based on this trend, I believe it necessary that any account of how scientific practice functions, viz. how problem-solving functions in science takes seriously the cognitive limitations of the human agents that perform the activities that make up scientific practice. As it happens, research on human problem-solving points to a crucial role for heuristics. This suggests that, as things stand, the notion of heuristics represents the best bet for elucidating the way scientific practice, understood as organized problem-solving activities, functions. The notion of heuristic has been part of the basic vocabulary in the psychology of reasoning since Kahneman, Slovic and Tversky's landmark study of heuristics and biases in the context of making decisions under conditions of uncertainty (1982). Some philosophers of science interested in providing accurate accounts of scientific practice have provided detailed arguments in support of the relevance of the notion of heuristics (Bechtel and Richardson 2010, chap. 1; Wimsatt 2007, chaps. 1–3; Hey 2014).

Early usage of the term "heuristic" referred not to specific strategies, but rather to a branch of study, which according to Polya aimed at studying the methods involved in discovery, adding that "Heuristic reasoning is reasoning not regarded as final and strict but provisional and plausible only, whose purpose is to discover the solution to the present problem" (Polya, 1957, 112-113). Over time, the term took on different, if related, meanings, from a particular kind of algorithm in artificial intelligence to cognitive strategies used by human and non-human animals alike to process information and make decisions efficiently. The next section attempts to clarify the concept of a heuristic and explain how such procedures may play a role in scientific practice.

3.2 Heuristics and their role in scientific practice

The difficulty with giving a definition of "heuristic" is that the term is employed in disciplines as varied as philosophy of mind, philosophy of science, psychology, cognitive science, legal theory, artificial intelligence and computer science (and likely a few more). Each discipline employs the term in ways that are *roughly* similar, but each has its own definition (Chow 2015, 977, 978). To make matters worse, authors within a given discipline may employ different definitions, and the stakes of debates over what, exactly, heuristics are go beyond terminological concerns. My aim here is to tease out features of heuristics that could constitute a satisfactory minimal definition, which in the present case is one that allows one to distinguish heuristics from other methods (i.e. avoids excessive generality) and is applicable to the analysis of scientific practice. Many debates concern the notion of heuristics as it is used in the context of providing an explanation for cognition in individual agents, which is not the topic of this dissertation. I will thus not delve too deeply in the debates around the precise nature of heuristics or what their role is in human cognition. It is sufficient for my purposes to show that the notion can be usefully applied in the context of philosophy of science (i.e. that it earns its explanatory keep) and that it can be defined in such a way as to avoid excessive generality (which would undermine its explanatory relevance). I am thus assuming that heuristics, or something like heuristics, play a key role in supporting human decision-making and problem-solving capacities. From this assumption (which is supported by the psychological literature), I claim that if one is to describe how scientific practice understood as problem-solving proceeds, then considering human cognitive capacities and skills is an obvious requirement.

An important thing to consider is that there is a negative connotation to the term "heuristic". As an example, consider Goodale and Milner's two-streams hypothesis, which holds that visual processing happens in two anatomically distinct areas of the brain and that this anatomical division of labor reflects a functional one (Goodale and Milner 1992; Goodale and Milner 2005; Goodale et al. 1991). Information processed in the dorsal stream, which projects from the primary visual cortex (V1) to the occipital and parietal lobes, concerns spatial visual information. It is involved in spatial localization of stimuli and in the regulation of movements (e.g. hand location and grip aperture in grasping behavior) based on visual information. It is also "unconscious", as the visual information it handles is not "consciously" available to the subject. The ventral stream, sometimes dubbed the "what" stream, projects from area V1 to the medial temporal lobe and is thought to handle object recognition, among other "conscious" forms of visual perception. This hypothesis was proposed following observations of effects of cortical lesions localized in patient DF's occipital cortex[23]. DF suffered from visual apperceptive agnosia, which is the inability to visually recognize objects, but had preserved motor capacities that relied on visual information. Common criticism of the hypothesis focuses on showing that the reality is more complex, arguing that the two-streams hypothesis is a "mere" heuristic: even though it gets something right, it grossly oversimplifies visual processing (see Schenk 2012 for an example). This kind of criticism is commonplace; branding something as a "mere" heuristic often indicates that the hypothesis or procedure is not a "real" one, but a mere shortcut.

This common perception of heuristics is not strictly speaking erroneous. The usual understanding of a heuristic is as an economical procedure that require less time to apply to problems than exhaustive methods, requires considering a reduced number of possible solutions, and employs "quick and dirty" rules of thumb to select a course of action, usually relying on a restricted set of cues to make a decision. Importantly, the economical character of heuristics is in part due to their lack of flexibility: they are faster and more economical because they are applied to each situation in the same way and do not undergo any sort of on-the-fly calibration to accommodate peculiarities of a given scenario. This means that heuristics are liable to systematic biases.

This risk of bias explains why heuristics are seen as "inferior" to, say, exhaustive algorithms that check every possible solution and rely on a complete analysis of the available information as opposed to a limited set of cues, at least in terms of accuracy and rigor. Thus, I want to make a very important clarification. The claim that scientists employ heuristics may be interpreted as derogative because of the intuition that any system that employs heuristics will be less accurate than a system that relies on more exhaustive decision procedures, in addition to being prone to systematic errors. A proponent of the traditional and idealized view of science discussed in chapter 1 might well take issue with characterizing scientific practice as relying on heuristic procedures, or at the very least would remark that if this is the state of scientific practice, then it needs

---

[23] Although the anatomical division had already been described by Ungerleider and Mishkin (1982).

to be changed to reflect ideal principles of inquiry. On the other hand, a proponent of practice-oriented approaches might be wary of characterizing crucial aspects of scientific inquiry in terms of heuristics because it amounts to an implicit criticism of scientific practice, one that may not be particularly productive. My goal in the following paragraphs is to show that if the idea that science is usefully characterized in terms of problem-solving is accepted, then looking at how humans (who are doing the science) solve problems as a starting point for understanding how scientists (who are also human) solve problems is a natural move. It does assume that scientists are not LaPlacean demons, but this assumption should not be interpreted as an implicit criticism of scientists. It is a realistic construal of their capabilities.

What are heuristics, exactly? Giving a good answer to that question will require some backtracking. Recall that on the constraint-inclusion model a problem consists in the demand for a certain goal state to be attained coupled with the set of all constraints on the goal state and the search for the goal state. Constraints eliminate possible solutions and thus understanding a problem requires understanding (some of) its constraints, at least those specific to a given problem (as some constraints might, in theory, apply to all problems). If it is the case that heuristic procedures are used in scientific practice, then evidence of that use will be found in the goal states and constraints that characterize scientific problems. Locating that evidence requires a clear explanation of how constraints work, that is, a clear understanding of how constraints interact with goal states and the process of attaining these goal states. To illustrate this and get a better understanding of *how* problems are solved, consider the example of a game of chess[24].

A game of chess is a well-defined problem-space: it includes a clear goal state, clear solution paths and expected outcomes, along with a clear set of rules corresponding to operators defining possible moves as well as path constraints (each board state permits a set number of moves). A solution is a sequence of moves (path) from a position P1 to a position Pn (checkmate). Thus, any solution to a game of chess will require a given number of steps (*n*) and at each step there are a set number of allowable moves (*M*). This

---

[24] This is a common example, but in this instance, it is adapted from Bechtel & Richardson 2010.

means that in any game of chess there are M^n possible paths from P1 to Pn, which is astronomical[25]. There are two things to note here. First, this example makes clear the extent to which *tractability* is an issue. Well-defined problems can be intractable if they have too many admissible solutions. Note that well-defined problems are a best-case scenario and yet may still present serious difficulties in terms of tractability. The problem of choosing which new refrigerator to purchase could well be intractable if one tries to take every possible factor into account instead of beginning with a pre-constrained search space (e.g. “I will not go out of town to buy a new refrigerator”). Some problems are ill-defined, meaning they lack a goal state or expected solutions, such that solving them will be very difficult even with the use of exhaustive algorithms. Furthermore, the example captures the sense in which a problem is defined not only by a demand for a goal state (checkmate) but also by the constraints imposed on the solution and the means by which one is to arrive at a solution (the rules of the game of chess). To push the example a bit further, consider the role that opening moves play in chess: it restricts the number of admissible paths to check-mate. Increasing the number of constraints on a problem by closing off possible solutions paths is a way to further constrain the search for a solution, albeit one that differs from the discovery of new constraints. Suppose a group of scientists is working on an explanatory problem. Trial-and-error approaches may reveal new constraints in the form of explanations that didn’t work (thus closing off possible paths to the goal of explaining a phenomenon), but it is also possible to “artificially” constrain a problem by applying a procedure designed to restrict the set of possible solutions without waiting for the solution space to become more constrained. Similarly, new constraints are “discovered” each turn that passes in a game of chess, because the board state changes and thus paths to checkmate become closed off, but an opening move imposes a constraint on the starting board state. This idea of imposing constraints on a problem summarizes the role heuristics play in problem-solving contexts[26].

---

[25] Decision-making problems encountered over the course of everyday life also pose issues for tractability. Optimizing the choice of a new television set is actually a very difficult problem, for instance.

[26] The analogy should not be taken too far: heuristics can be applied at any stage in scientific practice. The key idea is that they are strategies that impose additional constraints on a problem.

Nickles' characterization of problems takes up some important insights of Newell and Simon's (Simon 1977; Newell, Shaw, and Simon 1962), who stress the link between the constraints imposed on the search for solutions and heuristics[27]. The notion of heuristics has been extensively discussed in psychology (Kahneman, Slovic, and Tversky 1982; Gigerenzer 1999; Gilovich, Griffin, and Kahneman 2002; Gigerenzer and Brighton 2009) and philosophy (Chow 2015; Hey 2014; Wimsatt 2007, 2006; Bechtel and Richardson 2010) and can be understood as picking out procedures employed by cognitive agents to facilitate the solving of difficult problems. If, following Nickles, one takes seriously the idea of problem-solving as an apt description of the activities that make up scientific practice, then an adequate account of inquiry has to integrate the use of heuristics.

This idea is perhaps best articulated by Wimsatt (2007). On his view, philosophers of science should recognize that scientists rely on heuristics to develop theories, models, and experiments. Any view of science that does not take cognitive limitations seriously has no explanatory relevance. Attending to these limitations, in concert with a concern for descriptive accuracy, points to the central role of heuristics in science. Thus, Wimsatt holds that scientists employ error-prone methods that are heuristic in character[28]. Note that Wimsatt focuses on re-examining the foundations of philosophy of science (his goal is to outline a philosophy of science that takes seriously not only the cognitive limitations of scientists but also those of philosophers themselves). In contrast, my aim is to outline a descriptive framework that can be readily applied by philosophers of science. Thus, while my project is deeply indebted to Wimsatt's work, it is in pursuit of a different objective.

Going through the entire literature on heuristics would be a difficult task. Instead, I list some central features of heuristics below[29]. This list is not exhaustive and emphasizes

---

[27] In fact, Nickles' characterization of problems naturally points towards the importance of heuristics (Nickles 1981, n. 17).

[28] There is more to say about Wimsatt's view. Importantly, the heuristic character of scientific methods does not impair their usefulness, and it is possible to design meta-heuristics (rules for choosing which heuristic to use and how) to control for the biases that come with heuristic methods.

[29] These properties were selected because they appear in many accounts of heuristics. Moreover, the notion was originally developed in the context of accounting for problem solving in individual cognitive agents. In

characteristics of heuristics that are relevant in the context of analyzing scientific practice. That is, it omits properties related to questions of cognitive architecture and individual cognition. In a sense, the list was constructed by simplifying the set of available information about heuristics.

Table 5: Properties of heuristics

Fixed procedures; context-dependent.	(Gigerenzer 1999; Wimsatt 2007)
Fail in recognizable ways (fixed procedures entail bias).	(Gigerenzer 1999; Kahneman, Slovic, and Tversky 1982; Wimsatt 2007)
Facilitate problem solving by adding constraints on search for solution.	(Bechtel and Richardson 2010; Newell and Simon 1972)
"Transform" problem P1 into more constrained problem P2.	(McCauley 1986; Wimsatt 2007)
No guarantee of success.	(Wimsatt 2007; Darden 1991; Gigerenzer 1999; Gilovich, Griffin, and Kahneman 2002)

This needs to be unpacked. In the following sections I discuss each property and bring up relevant illustrative examples.

3.2.1 Heuristics are fixed procedures

First, heuristics are commonly seen as fixed procedures that are reliable in a given context. Often this is expressed in terms of fast and economical decision procedures that have a certain evolutionary history that made them useful at the time they evolved. The

---

the context I am interested in (scientific inquiry), some aspects will vary. For an informative review of the notion as it is used in psychology, see Chow (2015).

representativeness heuristic, for instance, is used to judge the probability of an event under uncertainty (Kahneman, Slovic, and Tversky 1982, 153–63). It evaluates the probability that a given event belongs to a certain category based on its similarity to other members of that category. This is a useful way to make a probability judgment in certain contexts (the large furry creature looks like another large furry creature that is dangerous and should be avoided)[30]. But in a different context it may lead to errors. A physician examines a patient presenting symptoms that resemble those of a rare hereditary disease. It so happens that a case of that disease was diagnosed the week before and because of this information the physician may diagnose the patient with said rare illness. This is a case of the *base rate neglect fallacy*[31]: since the rare disease has a very low rate of occurrence, it is more likely that the patient suffers from a common illness that shares symptoms with the rare one (or an atypical presentation of the former). The recent occurrence of an event can lead cognitive agents to overestimate the likelihood of another occurrence, even if the event is normally very rare. This might be advantageous in contexts where false positives have a lower cost than false negatives (e.g. predator avoidance), but in, say, medical contexts, false positives can have serious consequences, such as administering the wrong treatment. A more appropriate diagnostic principle is to consider the base rate of occurrence of a possible diagnosis. In other words: "when you hear hoofbeats, think of horses not zebras"[32].

3.2.2 Heuristics are liable to fail

This leads to the second property: heuristics tend to fail in recognizable, systematic ways. In the case of the representativeness heuristic, it fails by neglecting information related to

---

[30] There are pitfalls associated with making assumptions about the evolution of cognition and a "reconstructed" evolutionary environment. This example clearly highlights the fact a heuristic may be reliable in one context and lead to systematic error in another and should not be taken as more than an illustration.

[31] In the context of their discussion of normative debates in psychology, Bishop and Trout (2004, ch. 8) conduct a detailed analysis base rate neglect *qua* a systematic error in judgement. Interestingly, they discuss the possibility that such patterns as base rate neglect should not be characterized as errors. That said, so long as there is agreement on the fact that there are certain patterns of reasoning that can be identified, the usefulness of notions such as that of bias remains.

[32] Aphorism attributed to Theodore Woodward, M.D. (Sotos 1991, 1).

the base rate of an event occurring because it privileges surface information[33]. This kind of error can be tracked and reconstructed quite easily if one looks at the outcome and then for some a temporally proximate event that might have led the agent to make the decision they made. The representativeness heuristic may be very useful in survival contexts where the availability of information such as “another rabbit was attacked on route B recently” is likely to motivate the rabbit to use route A, even if it is longer than route B, because, again, *in this context false positives are not as costly as false negatives*. On the other hand, purchasing tornado insurance because there was a tornado not long ago in an area where tornadoes occur once every fifty years is a clearly irrational decision, but one that may be motivated by the application of the representativeness heuristic which in this case leads to the base-rate fallacy. If one knows how a heuristic works and what it does to modify a given problem, one can, to a degree, predict *how* the heuristic might fail.

These examples illustrate a common type of heuristic which excludes information from the search for a solution, privileging recent occurrences over historical information. This kind of informational restriction amounts to imposing a constraint to the problem, thereby further restricting the search space.

3.2.3 Heuristics facilitate problem-solving

Having a more constrained search space makes a problem more tractable because it has fewer possible solutions. This is the third property of heuristics: they facilitate problem solving by increasing constraints, which increases tractability. This sounds rather simple, but because of the way I have defined problems in this chapter and the previous one, it introduces a wrinkle which, as it turns out, is a crucial aspect of the model I develop here. That is to say heuristics *generate new problems.*

---

[33] From the perspective of a philosopher of science or a practitioner, predictable failure is a best-case scenario. More often than not, a foundational or otherwise important assumption or method won’t present the characteristics of a heuristic in an obvious manner. Scientists employing heuristics may not know whether they will fail. The important point is that they are *liable* to fail in recognizable and systematic ways, which lends normative potential to an approach based on the notion of heuristic. This is the source of Wimsatt’s suggestion that meta-heuristics can be designed to counteract biases: we can figure out how and when a heuristic will lead to error. It also provides the basis of the normative component of the model I sketch in later chapters.

Recall that on the definition of problems used here, problems are *constituted* by the constraints on the search for solutions. Modifying the constraints of a given problem will, technically, result in a new problem, one that is hopefully related to the original problem and whose solution is relevant to the aims of whomever was dealing with the original problem, but a new problem all the same. In this sense, heuristics do not, in and of themselves, *solve* problems. They *facilitate* the solving of problems, and the way they do this is by generating a new problem that is more tractable than the original one. This process may occur multiple times before a solution is arrived at, and it comes with serious liabilities. First, new problems may only be superficially related to the original problem. Second, heuristics may create, in addition to the "intended" new problem, various "unintended" problems, which may not be immediately apparent. I discuss such cases in later sections, but a quick example would be a case where an explanatory problem has constraints pertaining to admissible explanatory entities (e.g. excluding internal cognitive states from admissible components of explanations). This constraint might end up closing off important avenues for explanation, and this will not be immediately apparent unless experimental results are obtained that introduce some kind of anomaly or new phenomenon to be explained. There is rather more to say about this property.

3.2.4 Heuristics generate new problems

The generation of new problems is not just a liability. It is a crucial property of heuristics, one that makes it possible to build a model of scientific practice based on problem-solving. I discuss the implications of the problem-generation property in detail in section 3.3. For the purposes of this section, the main thing to take away here is that heuristics can generate the "wrong" problem, or generate unintended problems which may remain unacknowledged for long periods of time.

The possibility of generating the "wrong" problem, usually because the new constraints close off an otherwise correct solution, leads to the final property of heuristics, which is that they do not guarantee correct results.

3.2.5 Heuristics offer no guarantee of success

The new problems generated by heuristics may not be relevantly related to the original problem, or certain constraints might close off paths to the solution for the initial problem. A good illustration of such cases is illustrated in Wimsatt's discussion of reductionist heuristics (2007).

The category of reductionist heuristics comprises strategies directed at various aspects of inquiry, which are divided into subgroups by Wimsatt (conceptualization, model building/theory construction and experimental design). One heuristic of conceptualization, *descriptive localization*[34], provides an illustration of the discussion so far. This strategy consists in describing a relational property as if it were a monadic property. This makes a property that depends on both system and environment look like a property of the system only (static instead of relational). An example of this is fitness, which is a property of the relationship between organism and environment. A common heuristic is to keep the environment static, which makes it seem like one can treat fitness as a property of the organism[35]. Any problem related to fitness that is simplified in this way will result in an overly constrained search space. This is a common failure of heuristics: neglecting crucial information.

Such errors are not necessarily a strike against the use of heuristics (neither in general nor in science specifically). Errors can not only be corrected, but often are informative in their own right, as Bechtel and Richardson's (2010) discussion of decomposition and localization shows. Their discussion is conducted within the framework of mechanistic explanation, the basic idea being that explaining a given phenomenon amounts to identifying the parts and operations of the system that are responsible for it. Once the phenomena and the system are identified, explanatory work consists in identifying which parts of the system are responsible for the phenomena. To make this work possible, however, one must make the assumption that the system is *decomposable*. In effect, this

---

[34] Not to be confused with localization simpliciter, which is a different heuristic.

[35] Questions of fitness and adaptation become even more complex when the effect of organisms on the environment are factored in (see Barker 2007).

amounts to the assumption that whatever activity is performed by the system, it is performed linearly (or additively), with each part completing its own sub-function in turn (Bechtel and Richardson 2010, 24). Note, however, that most systems are only near-decomposable, if not entirely integrated (e.g. they rely on complex feedback loops and distributed processing, such that any one function cannot be localized to only one component part). In terms of problem-solving, attempting to provide a mechanistic explanation of a given set of phenomena without a prior decomposition of the system is an intractable problem. Decomposition will fail either because the system is not straightforwardly decomposable or because the resulting "picture" of the system is erroneous (there are other components that play important roles, etc.). Accordingly, the methodological value of decomposition claims lies not in their accuracy, but rather in their usefulness *qua* research heuristics. Decomposition is a fixed procedure that consists in splitting a system into discrete parts so that a problem involving the behavior of the system becomes a problem involving a compartmentalized system with identifiable parts. It is a different problem, but it is one that relies on a modified picture of the original system. The decomposition might be erroneous, but it constrains the search for an explanation or a description of the system, all the while creating potentially unrecognized and unintended problems that stem from it most likely being wrong.

Localization is the identification of the functions (or activities) specified in the decomposition stage to specific parts of the system. The grain of localization claims will depend on the grain of the decomposition. For instance, if the phenomenon of interest is a range of functions (such as "language production"), the localization claim may identify, e.g., Broca's area as a functional unit supporting language functions. If the phenomenon of interest is at a smaller scale, localization claims will focus on smaller functional units, and so on. Localization claims, just like decomposition claims, are often erroneous. For instance, early localizations of visual processing assigned area V1 as the sole brain region responsible for it, which turned out to be false (see Bechtel 2008, chap. 3 for a detailed account of how the visual system came to be understood). While both decomposition and localization often fail, they fail in recognizable and, in some cases, predictable ways. That is, there is a *type* of error that is likely to result from applying decomposition to a given system and being able to identify hypotheses and other claims *as* decomposition claims also points to potential errors that may impact the explanations or experimental results that result from frameworks built on such claims. The same goes for localization claims and, more generally, for any strategy that can be identified as a heuristic.

Note that because heuristics are liable to systematically fail, and because these failures will occur in recognizable or even predictable ways, analyzing practice in terms of the problems faced by scientists and the heuristic strategies they use to solve those problems opens the possibility of developing normative claims from the bottom-up as opposed to imposing them "from above". This idea is developed in more detail in chapters 5 and 6 but the gist of it is that in concert with an accurate description of science which enables the identification of the goals pursued by scientists, one can look at the strategies employed to solve the problems that crop up in the pursuit of these goals and then figure out the possible errors that these strategies might introduce. Of course, this is only possible after careful analysis of actual cases, and the focus on case studies and historical analysis is one of the defining features of the account I propose here.

3.3 Constraint-inclusion and heuristics: putting the pieces together

My aim for this section is to provide a preliminary statement of the model. So far, I have developed its two main components, viz. the constraint-inclusion model and the notion of heuristics, more or less independently. While the discussion of heuristics has drawn on the constraint-inclusion view of problems in a significant way, my goal here is to combine them into a descriptive model of scientific practice.

At the outset of chapter 2, I explained that the starting point for the model I am building is the claim that scientific practice can be understood as "organized or regulated activities aimed at the achievement of certain goals" ("SPSP" 2017). The precise meaning of "goals" in the definition is a matter of interpretation, and I believe there is at least one good pragmatic reason to adopt the view that the relevant meaning of "goals" will depend on *context*. That is, depending on the target of the philosophical analysis, the relevant goals may range from explanatory (identifying the mechanism supporting a phenomenon of interest) to investigative (such as the development of new experimental tools). There are many goal-oriented activities in science towards which philosophers might direct their analysis. Sensitivity to context of this kind makes it possible to avoid getting bogged down in debates about the "correct" level of analysis for understanding scientific practice. By "level of analysis", I mean the level at which philosophical analysis is conducted (e.g. the level of theory justification, or the level of experimental design, etc.). If there is such a thing, then a model based on this construal of scientific practice can be geared towards that particular level of analysis. In the rather more likely scenario that some form of pluralism about relevant levels of analysis turns out to be the most productive approach, then an analysis in terms of activities such as this one will be compatible with it.

The next step was to provide a characterization of "activities". Such a characterization had to be descriptively accurate and needed to be usable in the analysis of case studies and scientific practice in general. It also needed to meet four other conditions, but failure on the accuracy and in-practice applicability conditions was deemed sufficient to reject candidate characterizations (see chapter 1). In chapter 2 I provided some arguments to the conclusion that characterizing the activities that make up scientific practice as *problem-solving* activities seemed to at least have the potential to meet these conditions. Of course, saying that scientific practice is made up of problem-solving activities means there needs to be an account of problem-solving itself.

The first step towards such an account is a specification of what problems *are*. Thus, I reviewed three models of science as problem-solving: Laudan's theory of scientific progress (1978), Hattiangadi's theory of the historical and logical structure of problems (1978, 1979), and Nickles' constraint-inclusion model (1981). The first two failed to meet most of the six conditions, most notably the descriptive accuracy condition, because their definitions of "problem" excluded many activities that scientists routinely engage in. Nickles' model, on the other hand, offered not only the required descriptive breadth, but also promising connections to empirical work on human problem-solving capacities and skills. In fact, Nickles explicitly acknowledged the similarities between his account of problems and the work of Newell and Simon, who drew attention to the importance of heuristics in allowing for the economical and efficient solving or problems by human agents. Nickles' constraint-inclusion model defined a problem as the demand for a goal state to be reached combined with constraints on how to attain said goal state.

This definition of "problem" presents a crucial advantage: it allows the individuation of problems based on the *goals* pursued by scientists. That is, because a problem is defined by the goal state and the specifications on what the goal state should be and how it should be attained, it is possible to not only distinguish one problem from another (either because their goal states differ or because their constraints differ, while still being able to account for the way problems can be related to one another and how they might change over time or according to disciplinary context), but also to categorize problems based on these goal states. For instance, explanatory problems are problems for which the goal state is an explanation, while pragmatic problems are problems for which the goal state requires a practical workaround or the elaboration of a tool. The flexibility of the definition mirrors the flexibility of the construal of scientific practice as organized activities in the pursuit of a given goal: problems can be identified at various levels of analysis.

Because of the pre-existing links between Nickles' constraint-inclusion model and the psychological literature on problem-solving, the choice to tap into that body of work was a natural move. One of the adequacy conditions for a model of science as problem-solving was that the model should be compatible with a realistic view of scientists as human agents with limited cognitive resources and capacities, and the work on heuristics highlights their crucial role as economical procedures that allow humans to solve difficult problems rapidly and efficiently, while also providing robust evidence that human cognition does rely on this kind of procedure, at least to a significant degree. This, along with some other arguments to the conclusion that research on heuristics was relevant for my project, constituted the bulk of section 3.1. Section 3.2 provided a list of core features of heuristics along with an explication of each. I also discussed *how* heuristics function to facilitate problem-solving and decision-making, and then provided some examples of heuristics in science (decomposition, localization, and some reductionist heuristics).

At this point, all the elements required for a first formulation of the model have been made explicit. I emphasized, in section 3.2, that one property of heuristics would prove crucial to the elaboration of the model: the property of problem-generation. Since problems are defined by their constraints, and because heuristics facilitate problem-solving by adding constraints to problems and thus restricting the set of admissible solutions, what heuristics actually *do* is that they *create new problems*. The model I develop here takes this property as part of the core mechanic of scientific practice: it is through that process that research agendas develop, that new techniques and tools are designed, that unexpected conceptual inconsistencies crop up.

The descriptive model I propose conceptualizes scientific practice as a *dynamic* and *iterative* process. To make this clear, consider the following example, which is a "toy" version of the case study developed in chapter 4. Suppose scientists want to explain spatial learning in rodents. At a general level of analysis, the scientists have an explanatory goal, that of explaining spatial learning. They begin by observing the behavior of rats in a maze. The rats get better and better at finding food hidden within the maze. This pattern of behavior gives rise to a specific explanatory problem, that of providing an explanation for the way maze performance improves over repeated trials. The disciplinary context within which the scientists are conducting these experiments includes a strong behaviorist programme, such that they immediately reject possible solutions that include reference to internal cognitive states or representations. In other words, they impose the following constraint on their search for a solution: "the explanation must be couched in terms of observable behavior and stimulus-response interactions only", and so the set of possible solutions will be restricted to, say, explanations that rely on a stimulus-response account of learning that involves reinforcement as the main mechanism by which learning occurs[36]. From there, the application of this simplifying strategy generates new problems, such as the identification of relevant stimulus classes. The goal state of a problem of this sort, which involves

[36] Of course, this kind of procedure often happens implicitly, but the result is that the initial problem is transformed into a different problem that includes an additional constraint.

identifying the empirical and experimental correlates of theoretical entities invoked in the previous stage, will often be something like an experimental protocol, which specifies what experiments should be conducted along with an operationalization of the theoretical entities that are to be measured through the experiments. Such problems are often simplified by adding constraints derived from methodological precepts or theoretical commitments. This will result in the generation of still more problems related to the implementation of experimental protocols, often because of questions of availability of instruments, or because certain instruments have limitations on their use (e.g. using real objects inside a magnetic resonance imaging scanner (MRI) is difficult, and so one possible workaround is to substitute 3D objects with 2D images). Once experimental results are obtained, they are used to impose further constraints on the original explanatory problem, and the process starts over again.

Things can get rather more complicated. Imposing constraints on problems to transform them in more tractable problems enables scientists to make progress. That is, the way heuristics simplify problem-solving is also the way research agendas move from one problem to the next. However, it is also the way problems may become latent in theories and research agendas. Imagine that a group of researchers make the puzzling discovery that is that rats running mazes in which no food is hidden exhibit learning even in the absence of a reward. Specifically, while their error scores are high at the beginning, they drop precipitously once food is introduced in the maze, suggesting they have learned its layout prior to the food being introduced. This poses a relatively difficult problem, because it would appear from those results that learning can occur in the absence of reinforcement. This is an example of an *unintended* problem: it was always the case that because of the commitment to a strictly behaviorist explanation, the possibility of learning without reinforcement would create an explanatory problem, but the problem was not part of the spatial learning research agenda before those observations were made. What is happening here is the same thing as happens in normal cases when experimental results are used to transform an initial explanatory problem, only in this case the resulting problem is an explanatory problem for which there is no solution unless the constraint of excluding internal states from admissible explanations is removed[37]. More precisely, this is a case of an existing conceptual framework lacking appropriate resources (i.e., the behaviorist conceptual framework does not include the concept of internal representational states) to solve an explanatory problem. Solving it requires the adoption of a different conceptual framework or modifying an existing one.

What this example shows is that thinking about scientific practice in terms of problem-solving and problem-generation enables the construction of a descriptive narrative. Using historical analysis as a starting point, the constraint-inclusion model's notion of problem allows the identification of specific activities that scientists engaged in, and the notion of heuristics allows for an identification of the specific ways in which the problems faced by scientists were solved. Understanding the strategies used to conduct problem-solving activities can help make sense of why certain research avenues were pursued instead of others. Moreover, this analysis can be applied to problems other than explanatory problems, as I intend to show in later chapters.

Before moving on, I want to discuss Sullivan's analysis of the Morris water maze (2010). The Morris water maze is a commonly used experimental paradigm in neuroscience, which prompted Craver and Darden (2001) to use it as a case study for thinking about the ongoing discovery of the mechanism of spatial memory in neuroscience (Sullivan 2010, 262). The Morris water maze is an open field maze consisting of a pool filled with opaque water and placed in a room that contains a fixed set of visual cues placed outside the pool (distal cues). Rats placed in the pool will try to escape and swim around the pool. In the visible condition, a raised platform placed in one of the four quadrants of the pool is visible to the rat and its position remains constant across trials, although the rat's starting point varies randomly. The hidden condition is identical to the visible condition, except that the platform is silvery-white and placed just under the surface of the water such that it cannot be seen by the rodent. Training trials are performed during which the

[37] Again, this is a vastly oversimplified version of a real case. In reality, behaviorist researchers did find ways to accommodate those observations within their framework (Jensen 2006). Nonetheless this episode was crucial in prompting the adoption of representational frameworks by neuroscientists.

swim path and related variables (angles, length, swim time and time to escape, referred to as escape latency) are recorded. After training trials, the rodents are placed in a pool with no platform and the number of times it swims to the quadrant that held the platform, as well as how long it spends in that quadrant, are recorded.

Sullivan notes that there is no definitive answer to the question of what, precisely, is being investigated by researchers using the water maze: “spatial learning”, “spatial memory”, “Morris water maze performance” and other terms are put forward in the review literature. Despite this, the Morris water maze is taken to delineate a discrete phenomenon (i.e. “spatial memory”). Sullivan’s conclusion is that at best the Morris water maze circumscribes a discrete set of behavioral changes, not representational changes or changes in cognitive function. This is because, first, while Morris did develop a way to detect discrete changes in behavior, he did not specify what psychological function those changes in behavior indicated and, second, the Morris water maze became a popular tool in cellular and molecular neuroscience in a context where reducing cognitive functions to changes in behavior was the prevalent approach (Sullivan 2010, 275). That is, while researchers employed terms borrowed from disciplines that *were* interested in *what* the organisms learned (i.e. representational or cognitive changes), these terms, in the context of cognitive neurobiology, only referred to changes in response variables defined as observable changes in behavior (Sullivan 2010, 266-267). Specifically, the vast majority of researchers define learning as the acquisition of altered behavioral responses, which makes the use of terms that refer to cognitive or representational changes misleading. This can lead to errors down the line. For instance, since a change in behavior can be brought about by a variety of cognitive changes, using the discreteness of a set of behavioral changes to infer the discreteness of a set of representational changes can lead to grouping together heterogeneous representational processes or changes (Sullivan 2010, 268).

This analysis is very close to the one I conduct in the example above. On the problem-solving view, what is happening in the case of the Morris water maze is similar to what is happening when a behaviorist framework is adopted by researchers and closes off possible paths to a solution to an explanatory problem. The possibility of a mismatch between the discreteness of a set of behavioral changes and that of a set of representational or cognitive changes is a good example of an unintended problem that may remain undetected for long periods of time, and, as Sullivan claims the adherence to a behaviorist framework which may lead to problems like this one also resulted in investigators not knowing what exactly is under study in the water maze. This state of affairs can also be accounted for in terms of problem-solving and heuristics, by framing the lack of clarity with respect to what the water maze actually delineates as a conceptual problem that is a direct consequence of having adopted a behaviorist framework to solve a previous problem. Sullivan highlights that the fundamental problem is that even if investigators take the behavioral changes they detect to indicate specific changes in internal representations, those behavioral changes do not reveal much about what the content of the representational changes might be. That is, experimental paradigms like the water maze may lack the controls necessary for constraining what is learned by the organism, and thus lack the reliability required for discriminating between hypotheses about what the organisms actually learn (Sullivan 2010, 268). Again, this can be reformulated in terms of problem-solving: investigative contexts are rife with problems such as those identified in Sullivan's analysis, and researchers are always constrained in the way they are able to go about solving those problems. For instance, Sullivan notes that one way to solve the lack of necessary control in experimental paradigms is to conduct a detailed analysis of the features of learning paradigms and of what they afford in terms of what can be learned, which would not only be a time-consuming process, but also one that is unlikely to succeed in eliminating every possible confound.

I also want to say more about the iterative character of scientific practice as outlined in the example above. The importance of this iterative character has been noted in recent work by philosophers of science. Feest's analysis of the stabilization of phenomena is one such example. A key component of her analysis is the distinction between two senses of stabilization as it applies to phenomena. The first sense refers to the process "by which scientists empirically identify a given phenomenon or entity" (Feest 2011, 59), while the second sense refers to the gradual process by which the scientific community comes to agree that a given phenomenon is in fact a robust feature of the world and not an experimental or instrumental artefact (Feest 2011, 59). Feest argues that these two notions of stabilization can be understood as applying to two different kinds of phenomena: the first notion applies to empirical data patterns found in the world or produced in the laboratory (surface regularities), while the second notion applies to underlying, or hidden regularities (Feest 2011, 63). The crucial point of Feest's analysis is that to be able to empirically investigate claims about the stability or robustness of phenomena (that is, to stabilize phenomena in the second sense), we first need to assume that it is possible to identify the phenomena in question by, for instance, using an instrument that produces data we take as instantiating surface phenomena which are treated as indicating the (hidden) phenomena. In other words, our ontology of surface phenomena informs our ontology of hidden phenomena. On the other hand, trying to validate claims about hidden phenomena may result in the realization that the surface phenomena used as indicators of hidden phenomena are not adequate for the task (Feest 2011, 66). Feest's point is that surface regularities are necessary to the investigation of hidden regularities, but that the results of that investigation may change how the surface regularities are classified (Feest 2011, 70). This highlights the dynamical and iterative character of the process of stabilizing phenomena[38]. The model I propose here hinges on a similar claim about scientific practice in general and essentially extends this insight beyond the specific problems associated with stabilizing phenomena[39].

The core insight of this chapter is that the property of heuristics of generating new problems points to an informationally rich process that is ripe for analysis. It is in that process that, with the benefit of hindsight, a philosopher of science can identify branching explanatory paths that ended up not being explored until later (if at all), locate the point at which an erroneous assumption or conceptual framework was adopted or locate the origin of latent problems that crop up later in the historical development of a

---

[38] Importantly, the idea that the scientific process is in some way iterative is not new. Popper's understanding of science, for instance, emphasizes the iterative cycle of conjecture and refutation. What is novel about approaches such as Feest's, Sullivan's, and mine, is that the iterative process described includes a wider range of scientific practices or emphasizes different aspects of those practices.

[39] Other philosophers of science have insisted on the importance of appreciating the iterative character of various activities involved in investigative practice. Sullivan (2016) points out that the iterative process of explicating and validating experimental constructs should be part of the experimental process.

research agenda or in the context of interfield reduction or unification. For instance, debates pertaining to the difficulty of reconciling philosophical notions of representation with notions of neural representation can be clarified and potentially dissolved if one understands why neuroscientists made the decision of adopting representational frameworks in the first place. In this case, the incompatibility of the two kinds of notions of representation is a conceptual problem generated by the recourse to representational frameworks in neuroscience, but one that remained latent until philosophers began showing interest for the study of the brain. Moreover, it arguably is a problem of doubtful significance to anyone save philosophers: the notion of neural representation serves a specific function in neuroscience regardless of how compatible it is with philosophical notions of representation[40].

Another important feature of the model not made explicit in the example above is that it allows for the specification of a taxonomy of scientific problems. The example presented an iterative process, from an initial explanatory problem to the pragmatic problems involved in the implementation of experimental protocols. Recall that the notion of problem at work here allows for problems to be individuated based on their goal states, and these goal states can be grouped in categories. A first approximation of this taxonomy is presented in table 2 below. It should be noted that these categories are quite broad, and that the model I outline is preliminary and probably does not capture all aspects of scientific practice. This being said, the taxonomy of scientific problems is intended as a rough guide for characterizing problems and for thinking about the relationship between problems in the context of the development of a research agenda.

Table 6: Taxonomy of scientific problems

	Conceptual problems	**Empirical problems**	**Practical problems**

[40] This is not meant to imply that there is no value in studying the concept of representation as it is used in the neurosciences. The concept plays important roles in many areas (see Sullivan 2010 for an analysis of its role in cognitive neurobiology, Bechtel 2014 for an analysis of its role in cognitive neuroscience, among others).

Typical context	Concern basic framework of inquiry.	Concern physical substrate of theoretical terms involved in frameworks proposed at conceptual level.	Concern implementation of empirical inquiry (controlling for confounds, operationalizing various variables, etc.)
Typical goal state	Explanatory hypothesis that can direct inquiry, conceptual framework, (e.g. behaviorist framework).	Identification of the (admissible) empirical indicators of the components of the conceptual framework.	Experimental paradigm, instrument development.
Relationship to other problems	Solving conceptual problems introduces other problems, some of which are empirical. Conceptual problems can also derive from experimental results, which are influenced by solutions to practical problems.	Derive from a proposed solution to a conceptual problem. Solving them will generate practical problems.	Solutions to practical problems determine how heuristics and frameworks are evaluated and revised. Because solutions to practical problems influence experimental results.
Example	Example: explanatory problems.	Example: identifying relevant stimuli classes for spatial learning.	Example: limitations of existing experimental tools.

If it is possible to categorize problems in this way, it should also be possible to categorize heuristics in a similar manner. Heuristics can be grouped on the basis of the manner in which they add constraints to a problem, or the kinds of constrains likely to result from the application of a given heuristics. They can also be categorized based on the kind of problem they are commonly used to solve. Localization, for instance, is a heuristic that is more likely to be useful when dealing with empirical problems: it imposes a specific organization on a physical system to as to allow for the empirical investigation of that system.

It is important to flag that the details of the model as presented here are provisional. Likely, it will need to be modified following the application of the model to the case study in chapter 4. The presentation of the model in this section is meant as a proof of concept, a demonstration of the kind of work that can be accomplished when the constraint-inclusion model is put together with the notion of heuristic as developed in sections 3.1 and 3.2. Figure 1 below is a representation of the model at this preliminary stage.

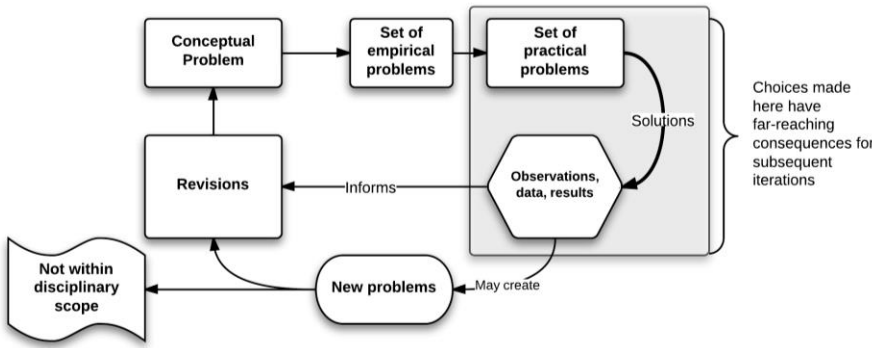


Figure 3: The dynamic-iterative model of scientific practice

Before moving on to the conclusion of this chapter, I want to draw attention to normative potential of the dynamic-iterative model. One important feature of heuristics is that they are fixed procedures: by and large they are applied to problems in the same way every time, and the constraints a given heuristic imposes on a problem will be the same kind of constraints each time. Their usefulness is derived from this lack of flexibility: it is because heuristics are applied without modification to various problems that they present an advantage over exhaustive algorithms. It is also because of this that heuristics are biased. They can, and do, fail in systematic ways. For instance, the representativeness heuristic leads to at least two specific errors: base rate neglect and the conjunction fallacy, an error that occurs when specific conditions are assumed to be more probably than a single general condition (Kahneman, Slovic, and Tversky 1982)[41]. The notion of

[41] This fallacy is thought to result from applying the representativeness heuristic to assess the probability of an event or state of the world. The classic formulation of the effect is derived from studies using the so-called “Linda problem”. Participants are presented with some information about a fictional person, Linda: she has a degree in philosophy, she is involved in social causes, and she works at a bank. Participants are then asked to compare the likelihood of two statements: “Linda is a feminist and a bank teller” and “Linda

heuristic is not merely convenient as a conceptual tool that is compatible with the constraint-inclusion model and allows for potentially accurate descriptions of scientific practice. By their very nature, heuristics can be studied and catalogued, such that it is possible to provide normative recommendations based on the specific kinds of errors that a given heuristic is liable to introduce. This is a rather different kind of normativity than the one proponents of the Practice Turn criticized, as it is grounded in an accurate description of scientific practice first and foremost. Moreover, that heuristics can introduce errors, and what those errors are, is not by itself sufficient for a relevant normative program. Such a program also requires an accurate conceptualization of the goals of scientists, and it is only relative to those goals that errors can be assessed in terms of how serious they would be, and how they can be avoided. Thus, the model I propose has the potential for normative contributions that can be directly relevant to scientists.

3.4 Conclusion

In summary, the dynamic-iterative model of scientific practice starts from the claim that scientific practice can be construed as made up of numerous instances of problem-solving activities. The analysis of these problem-solving activities relies on two core components: the constraint-inclusion view of problems and the related commitment to the role of heuristics in the problem-solving activities that make up scientific practice. Together, these components allow for the construction of descriptive narratives for case studies, and for the elaboration of conceptual tools like a taxonomy of scientific problems as well as a taxonomy of scientific heuristics. The narratives are constructed in such a way as to make clear the specific problems scientists faced, the strategies they used to constrain those

---

is a bank teller”. In their original reporting of the effect, Kahneman, Slovic and Tversky (1982) noted that most participants judged that the first statement (the conjunction) was more likely, which is false since the conjunction requires both conjunctions to be true, while the second statement only requires Linda to be a bank teller. They take this result to show that humans rely on the heuristic of representativeness to assess probability: the short biography of Linda makes it seems likely that she would be a feminist, and the conjunctive statement is thus more *representative*. This interpretation has been criticized (notably in Gigerenzer 1996), but nevertheless the example gets the point across.

problems and transform them into integral parts of their research agendas, and how the observations that result from the iterative process modify research questions and problems. The model thus relies on historical analysis as a starting point.

Recall that my aim is to construct a model that can provide an accurate description of scientific practice, that is in-practice applicable to actual cases, and that allows for the possibility of making relevant normative contributions to science. The foregoing only provides *theoretical* support for the claim that the dynamic-iterative model meets these requirements. I believe that the only way to convincingly support the claim is to engage in an actual application of the model. Thus, the next chapter introduces the dissertation's central case study, viz. the case of place cell research. The chapter will function as an illustration of the descriptive power of the model in addition to demonstrating the philosophical relevance of it, as it provides insight into a long-standing philosophical debate about the status of neural representations.

Chapter 4

4 Applying the dynamic-iterative model

One of the most important aims of this project is the in-practice applicability and usefulness of the model it outlines. Moreover, I have rejected other problem-solving approaches in part *because* they lacked such in-practice applicability (see chapter 2). It is thus crucial that I demonstrate that my model can be fruitfully applied to real cases. To this end, this chapter focuses on the case of place cells. While research on place cells and their function is ongoing, their role in spatial learning and navigation is relatively well-understood. A further motivation for this choice of case study is that place cell research directly connects to philosophical debates about the use of representational frameworks in neuroscience and the status of neural representations and the continuity (or lack thereof) between philosophical and scientific notions of representation. Moreover, because the use of representational frameworks in neuroscience is a common research strategy with a long history, it is easy to find cases in which the strategy breaks down, making it possible to showcase the normative potential of the dynamic-iterative model.

The chapter is divided in three sections. Section 4.1 introduces the case of place cells and is itself divided in two sub-sections. Section 4.1.1 discusses the conceptual background against which O'Keefe and his colleagues developed the theory of the hippocampus as a cognitive mapping system supported by the activity of place cells. Section 4.1.2 clarifies the sense in which neurons can be said to represent, and how the representational view of neural activity translates into research questions in neuroscience. Section 4.2 consists in a summary of Bechtel's (2014) analysis of the case of place cells. I identify a lacuna in Bechtel's analysis, and conduct my own analysis of the case in section 4.3, using the dynamic-iterative model to supplement Bechtel's analysis.

To anticipate on chapter 5, one of the main contributions of my dissertation is providing a framework that enables philosophers of science to recognize the heuristic nature of some key research postulates, while providing a rationale for why such heuristic postulates are not to be dismissed as “merely” heuristic. Getting clear on the status and function of the notion of representation is a necessary step towards that goal.

4.1 Place cells and neural representation

The goal of this section is to formulate a clear notion of neural representation that can be used in the analysis of the case of place cells. To achieve this, I begin by providing a short survey of place cells, focusing on the discovery and subsequent corroboration of their role in spatial learning and navigation. Neural representation comes into play when trying to elucidate how neuroscientists construe the activity of place cells, both in terms of *what* they do and *how* they do it. After providing an account of neural representation, I discuss some objections to the use of representational notions from philosophy of mind, neuroscience and cognitive science.

4.1.1 Place Cells: The Conceptual Background

Place cells were discovered by O’Keefe and Dostrovsky (1971) while conducting single-cell recording in the rat hippocampus while the animals were running around an open arena. They noticed that some cells were active only when the rat was in a specific place – rats were manually situated and held in place. The location of the electrodes was not selected randomly, as the role of the hippocampus in memory was known for some time before O’Keefe and Dostrovsky’s experiment. Scoville and Milner’s study of Henry Molaison (then known as H.M), whose drug-resistant epilepsy required surgical intervention, had shown that the hippocampus played an important in new memory formation (Scoville and Milner 1957). H.M. had undergone bilateral medial temporal lobectomy, essentially removing most of his hippocampus and surrounding cortex on both sides. While he could remember details from his past, he was affected with severe anterograde amnesia and was thus unable to form new memories (although his short-term and procedural memories seemed to remain functional). Based on the findings of the 1971 animal study, O’Keefe and Dostrovsky suggested that cells in CA1 of the hippocampus provided the brain with a spatial map of the environment. In a subsequent study, O’Keefe modified the experimental setup from manually situating the rats to allowing them to run freely in a three-arm maze, the sides of which were open (1976).

Still recording from the same region, he noticed that some cells (26 out of 50 recorded cells) responded to a specific location. On this basis, he asserted that the firing of certain neurons provided a map of the rat's location in the world.

In a way, the discovery of place cells vindicated the work of psychologist Edward Tolman, who in 1948 had suggested that the performance of rats in spatial problem-solving tasks could be explained by postulating that rats built a map-like representation of their environment. Tolman's proposal came during the height of behaviorism, and his view was opposed to stimulus-response views such as Hull's (1952), which remained dominant until the emergence of cognitive psychology in the 1960's. On the other hand, the discovery of place cells raised many questions, some of which had already cropped up in Tolman's work: how do these maps function? What kind of maps are they (comprehensive or linear, path-based)? What kind of stimuli is required to allow the construction of such maps? In addition, O'Keefe and Dostrovsky essentially brought the brain in the picture, and along with it several empirical questions related to the way place cells work to provide a map of the environment: how are those maps constructed in terms of neural activity? Where do place cells get their inputs from?

In the coming years O'Keefe would split his time between pursuing empirical investigation into these questions (some of which are reviewed in section 4.3) and developing an ambitious theory of spatial navigation, one that conceived of the hippocampus as a cognitive map. In 1978 O'Keefe and Nadel published a book that outlined that theory, and understanding this theory is crucial to understanding the connection between place cell research and the philosophical issues surrounding neural representation. It also provides useful background for the upcoming discussion of how place cells function. The book's first chapter is a detailed discussion of the philosophical roots of the concepts of space and spatial perception, covering an impressive range of positions including Kant, Leibniz, Newton, Poincaré and Piaget to name a few. O'Keefe and Nadel draw a distinction between two conceptions of space: relative space (egocentric) and absolute space (viewer-independent). Note that on their view the components of space are *places* (O'Keefe and Nadel 1978, 86). They go on to argue that absolute (or unitary) space is a more promising basis for constructing a theory of spatial perception, in part because models that attempt to start with egocentric space are unable to account for crucial features of spatial navigation:

> Most contemporary models start with relative space and attempt to build first a framework and then metrical space. These models rarely get off the ground, stumbling over the problems of re-identification and movement. It seems reasonable to conclude, first, that there is a clear need for the concept of unitary space. (O'Keefe and Nadel 1978, 59)

They add that Kant was likely correct and that absolute space could only be innate:

> Further, it appears that this framework cannot be acquired through experience; it must be available soon after birth, for the processes of localization, identification and the coherent organization of experience depend on it. (O'Keefe and Nadel 1978, 59)

Of course, O'Keefe and Nadel, being neuroscientists, were interested in the neural basis of something like an innate understanding of absolute space and, based in part on O'Keefe's work on place cells, suggested that the hippocampus was a likely candidate for this. Before discussing this, I believe it worthwhile to spend some time clarifying the conceptual background against which they developed their theory of the hippocampus as a cognitive map. Much of this background can be gleaned from the second chapter of the book.

The chapter focuses on specifying types of behavior that would require something like cognitive mapping, that is, spatial behavior that cannot be explained without invoking the notion of an internal map of absolute space. This is done through a review of the literature on navigation in birds and humans and place learning in rats. Homing and migration in birds had been a topic of interest for some time and O'Keefe and Nadel note that early studies of homing in the noddy and sooty tern by Watson and Lashley (1915) had mostly negative results. While they could show that birds were able to return to their nesting ground on Key Island, Florida when released as far as 1000 miles from it, they did not manage to provide support for any particular hypothesis as to how the birds could manage this (O'Keefe and Nadel 1978, 63). Most hypotheses at the time postulated that the animals relied on a set of stimuli to guide them towards their destination or referred to hereditary memory, and no evidence in support of either type of hypothesis was to be found. They were more successful in accounting for nest-finding abilities, showing that birds relied on specific routes made up of landmarks, which did not require much in the way of what they called "ideational processes", viz. cognitive processes more complex than habit-formed reactions to prominent visual stimuli. Watson and Lashley knew that their analysis of nest-finding could not be extended to long-distance homing (in part because they had calculated that the visual stimuli used in nest-finding was not available to birds in long-distance homing), but they did not think to explore the possibility that these so-called ideational processes might be involved in the latter. With attempts to identify the crucial stimulus class involved in homing failing systematically, research eventually turned to more complex cognitive models. For instance, Griffin (1955), recognized that animals could navigate in at least three ways: piloting (steering by landmarks), compass steering (heading in a constant compass direction) and true navigation (heading towards a goal regardless of starting point and direction required to reach the goal). O'Keefe and Nadel note that these three navigation strategies roughly correspond to what they call guidance, orientation and map-following, respectively (1978, 64). True homing behavior relies on the third type of strategy, since it is the only one that does not rely on landmarks (in the way that nest-finding does, for instance) and that is flexible enough to allow birds to return to their nest from diverse locations and over long distances. "True navigation" is best explained by the map-and-compass hypothesis, which postulates that birds have a map which they can use to locate their starting and ending points as well as their own location, and some form of internal compass, which allows them to calculate the direction they should head towards to reach their goal (O'Keefe and Nadel 1978, 65). At that point, research had been mostly focused on understanding the features of the compass, and identified three robust mechanisms used by birds to compute compass direction (sun position and internal biological clock, geomagnetism and in migratory birds who fly by night, star maps). While little attention had been paid to the concept of a cognitive map, O'Keefe and Nadel were confident that such maps were necessary to explain the homing capacities of birds:

> [The] birds are not using landmarks of any sort, nor are they following routes. Airplane tracking of individual birds shows that they seldom, if ever, return home by the same route from the same release site and that even when, on the later returns of a series, they come across landmarks associated with previous return flights they do not change flight plans and follow the previously used route. (O'Keefe and Nadel 1978, 66)

If the behavior exhibited by birds cannot be explained by the use of landmarks or routes[42], then the only option left is that they employ internal, map-like representations. O'Keefe and Nadel also briefly bring up Gladwin's (1970) descriptions of the navigational techniques employed by South Seas islanders (Puluwatans) which seem to mirror the navigational devices discussed by Griffin (1955): landmarks to guide the navigator at the end of the journey, compass steering aided by the sun and stars to maintain course and, crucially, cognitive mapping of the region (O'Keefe and Nadel 1978, 66). These maps locate each island relative to each other along with various seamarks and, importantly, a superimposed star map made up of 32 prominent stars, chosen in such a way as to practically cover the 360º of the compass. It seems unlikely that O'Keefe and Nadel take the maps employed by the Puluwatans to be the *same* maps employed in the case of homing. Aside from probable inter-species differences in how a cognitive map might be implemented in the brain, the Puluwatan maps are likely to be, in part, socially constructed and communicated in an explicit manner that doesn't happen in the case of birds. What *is* relevantly similar between birds and Puluwatan sailors, however, is that they accomplish complex spatial problem-solving in functionally similar

---

[42] I discuss the distinction between routes and maps below, but the former is essentially a sequential set of landmarks along with specifications for how to respond to each landmark whereas the latter is an aggregate of information which does not include any specifications on responses (see O'Keefe and Nadel 1978, 63).

ways: both rely on 1) a sophisticated mapping device and 2) a set of devices for operating on that mapping system, allowing the organism to navigate its environment.

To sum up, the point O'Keefe and Nadel are trying to make is that any explanation of complex spatial behavior *requires* a notion of cognitive mapping. Further evidence for this claim is drawn from the literature on place learning, a learning process thought to be distinct from the commonly accepted stimulus-response accounts of learning characteristic of the behaviorist research program. O'Keefe and Nadel bring up Tolman's well-known 1948 article, which reviewed the evidence in support of his hypothesis that rats used what he called "field maps" (which are, for all intents and purposes, akin to cognitive maps) of their environment to get from one point to another. Since I discuss Tolman's work in detail in the next section, a short summary will suffice here. One important study that provided support for field theory was conducted by Tolman, Ritchie and Kalish in 1946. Rats were put in a training maze and were trained to follow path BCDEF from start point A to goal G.

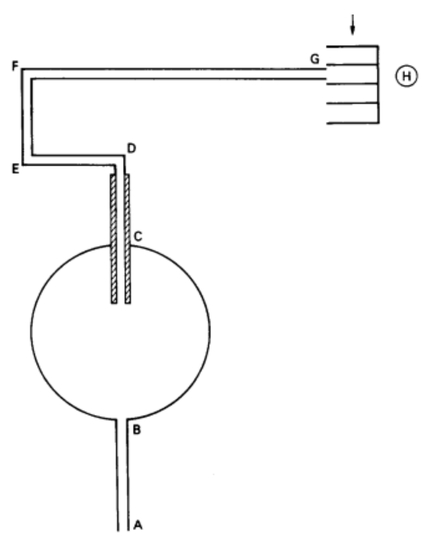


Figure 4: Maze used in preliminary training in Tolman et al. 1946. The "H" represents the general direction of the goal box. (Image taken from O'Keefe and Nadel 1978, 70)

After training, the rats were put in a sunburst maze and allowed to choose which path to follow to get to their goal. Most rats ended up choosing the arm that pointed directly in the direction of the goal (represented by the circled "H" in figures 4 and 5). They exhibited what O'Keefe and Nadel call "classic detour behavior, the animal going to a *place rather than making a particular response*" (O'Keefe and Nadel 1978, 71, emphasis added).

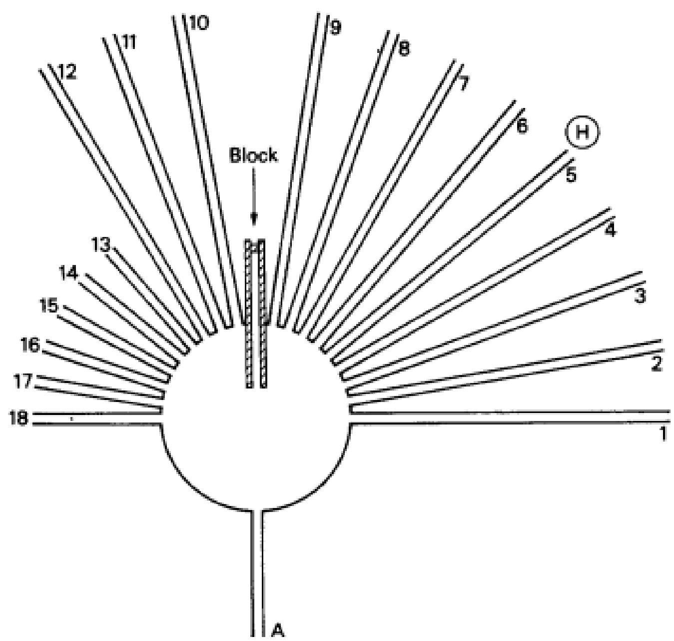


Figure 5: Sunburst maze used in Tolman et al. 1946. Most rats headed in the direction H, which is where the goal box was located during training. (Image taken from O'Keefe and Nadel 1978, 71)

Hebb (1938a, 1938b, 1949) described similar behavior, such as rats taught to run to a food dish placed at the edge of an open table running to the *place* where the food was after the table and food were rotated 90º relative to the room, or approaching goals that are located in the correct *place* but containing incorrect cues (or different cues from the ones present during training). These results seemed to indicate that rats could specifically learn *places* as opposed to stimulus-dependent responses, a capacity described by Tryon as

> the *native* capacity of the animal to evolve directional abstractions regarding the plan of the maze. *These abstractions are developed out of sensations derived from*

> *stimuli received from the maze during learning* … when a rat has developed these directional sets, they guide his movements in the maze, even in the presence of radical stimulus changes—he becomes free of the specific stimulus features of the maze. (Tryon 1939, 414, in O'Keefe and Nadel 1978, 72, emphasis in original)

While Hebb agreed with the overall characterization of this capacity, he thought that information from the rat's distal environment was more important than proximal stimuli in defining places and enabling orientation (1938a). This focus on the importance of distal cues would eventually undermine the stimulus-response (S-R) account of place learning phenomena and maze behavior. What happened was that proponents of the S-R view denied that there was anything special going on and that place learning was simply another instance of S-R learning (e.g. Restle 1957). O'Keefe and Nadel point out that the S-R view left unexplained the apparent flexibility of place learning, as well as the apparent effectiveness of distal cues in enabling the specification of places. On this last point, Hebb (1949) suggested distal cues were useful because of their constancy compared to proximal stimuli, but O'Keefe and Nadel point out that this was perhaps the wrong way to think about it, because distal cues cannot enable an organism to distinguish between two places by themselves. Rather, their importance was more likely tied to the specification of *direction* since they could be used in an analog manner as the stars in the Puluwatan sailor's cognitive map. Moreover, they doubt that distal cues can enable a process such as place learning since "it is well known that when a cue is spatially distant from the site of the response to be made to it organisms have great difficulty in learning (cf. Cowey 1968)" (O'Keefe and Nadel 1978, 74-75), especially given that place learning seemed to occur much faster than cue learning (75).

This overview of the literature on place learning essentially supplements the claim O'Keefe and Nadel made when discussing complex spatial problem-solving. Not only are spatial capacities necessarily supported by devices operating on map-like representations of space, but the learning mechanisms associated with the development of spatial capacities appears to be a specific kind of learning, one that does not rely on associative processes pairing stimuli and responses[43]. Before moving on to questions of anatomy and physiology, I want to discuss the distinction between *routes* and *maps*. This is important because while O'Keefe and Nadel think of what the hippocampus does in terms of mapping, I have not yet given a clear definition of what a map would be in this context.

One problem with Tolman's proposal is that it included no specifications on the properties of cognitive maps nor any comparison between maps and other means of navigating in the world (O'Keefe and Nadel 1978, 80). O'Keefe and Nadel begin by stating that there are two broad types of strategies for getting from one point to another. *Routes* are lists of stimulus-response-stimulus (S-R-S) instructions or commands that lead an organism from stimulus to stimulus until it reaches its goal. The stimuli can be visual (landmarks), tactile ("continue until you're walking on softer ground") or tied to any other sensory modality (although some would likely be far less effective than others depending on species and context). Moreover, routes work by providing either *guidance* or *orientation*, the former being tied to stimuli or cues (in that the cues acts as *guides*) while the latter is tied to responses (in that they emphasize the response requirement while neglecting cues). Orientations can be understood as instructions for maintaining the organism's body in a certain direction and are often aided by the inclusion of some reference point that is to be kept in a specific region of, say, the visual field, since using only interoceptive and proprioceptive cues to maintain a straight path is difficult (O'Keefe and Nadel 1978, 82). In both cases, routes rely on egocentric space, since they are comprised of sequential responses to specific stimuli. While this makes routes very simple to use and forgiving in terms of cognitive demands, their reliance on sequence and egocentric space makes them inflexible, which in turn makes them very vulnerable to landmark or mental degradation. If a landmark disappears or changes significantly, the entirety of the route may be rendered useless. Similarly, errors made in following a route

[43] It seems relevant to flag that the lack of a detailed discussion of latent learning strange in this context, as Tolman's 1948 paper heavily relied on evidence that rats could learn the spatial layout of their environment *without* the presence of rewards to undermine the S-R account of spatial learning. Thus, the evidence cited by Tolman also supported O'Keefe and Nadel's contention that map-like representations of space were a necessary explanatory resource in this context.

can be hard to recover from, since the route provides no information outside of the sequence of responses that must be followed to reach the goal[44].

Maps, on the other hand, offer a much more flexible way to get around. O'Keefe and Nadel define maps as representations of parts of space, adding that space being constituted of places, a map can further be defined as a "representation of a set of connected places which are systematically related to each other by a group of spatial transformation rules" (1978, 86). Note that on their view space and place are understood as what they call "conceptual primitives" which are not reducible to other entities. Space is not defined in terms of the relationships between the objects occupying it, and the habit of locating a place by referring to the object occupying it (e.g. "Kramer left the cantaloupe where the air conditioner is now") is mere convenience. Maps are further divided into topographic and thematic maps, and O'Keefe and Nadel are interested in the former, which uses symbols to represent objects and places in space as well as system that roughly preserves Euclidean principles (1978, 86). Maps are nothing over and above representations of space understood as sets of places and symbols representing objects located in space. Such representations include no specifications on sequences of responses to follow to get to a goal, but they do provide information an organism might use to reach a variety of different places, or to decide on a location for a nest, or to avoid a dangerous area. In effect, the main advantage of maps is that they are highly flexible, enabling one to pursue a variety of goals in a variety of ways. They can be informationally dense or simple and, importantly, are safer than routes because they are more resistant to mental and landmark degradation. If a landmark is nowhere to be found, one can self-locate by using some other landmark included in the map; they can be updated in a way that routes cannot. Similarly, errors made while following a route are easier to correct with a map. Of course, this means that maps are computationally more demanding than routes and the requirement for encoding of the information into sets of

[44] Interestingly, this mirrors some features of heuristics: much like routes, heuristics are forgiving in terms of temporal and cognitive demands, but can break down when their context of use is altered.

symbols makes them harder to use and learn than simple routes (O'Keefe and Nadel 1978, 86-89).

Of course, effective spatial navigation requires both maps and routes. Maps provide a framework within which routes can be designed and updated. O'Keefe and Nadel's key assertion is that the hippocampus acts as a cognitive mapping system, which they call the *locale* system, while the route, or *taxon,* system is handled somewhere else in the brain (their focus being on the locale system, they do not discuss the implementation of the taxon system, only what is necessary for them to make predictions as to the behavioral consequences of hippocampal lesions, see O'Keefe and Nadel, 1978, 90). Accordingly, a place in the environment is defined in terms of the activation of a specific group of hippocampal neurons. O'Keefe and Nadel bring up the notion of *place representation* here:

> this can be taken as a part of a cognitive map, while conversely a map can be viewed as a set of ordered, connected places. Such a place representation can be activated in either of two ways: (1) externally, by the simultaneous occurrence of two or more sensory inputs with the appropriate spatial co-ordinates in egocentric space; (2) internally, by an input from another place representation coupled with a signal from the motor system concerning the magnitude and orientation of a movement. (O'Keefe and Nadel 1978, 93)

The first mode of activation relies on complex sets of integrated cues rather than a single cue or class of cues such that removing one or more will not prevent whatever remains from uniquely specifying a given place (so long as enough remain). This feature is important, because it is part of what distinguishes place-learning from simple cue-learning (O'Keefe and Nadel 1978, 93). The second mode of activation is used in conditions where the animal might need to compute the expected changes in its environment. The cognitive mapping system is also assumed to contain a map for each environment experienced by the animal. The construction of the maps is thought to begin with the initial assignment of a place representation to a specific location in a new environment. From this initial assignment, the rest of the place representations are progressively assigned to other locations in a way that is consistent with their relationship to the initial representation. The exploratory behavior observed in rats is thus, on this view, designed to build and update cognitive maps. Once the map has been built, the animal can use it to solve various problems: it can label certain places as safe or dangerous, compute routes between two or more points, avoid or seek certain locations, etc. O'Keefe and Nadel take this to explain the behavior observed in maze studies such as Tolman's, as well as the capacity of rats to behave in a consistent manner despite small changes in their environment (1978, 95).

The conceptual background of place cell research is built up from theoretical conclusions about the kinds of mechanisms that are thought to be necessary to support complex spatial behavior, with supporting evidence drawn from studies of homing and navigation in birds and humans and place-learning in rats. Beginning with an epistemological argument to the conclusion that accounting for the capacity to understand and use absolute space was the necessary first step in explaining spatial problem-solving capacities, O'Keefe and Nadel argued that any account of such capacities required the notion of cognitive mapping. The notion needed to be further defined in a way that made it applicable to the context of neurological explanations, and so O'Keefe and Nadel drew a clear distinction between maps and routes, teasing out those features of maps that could be included in an operational definition of cognitive mapping.

Of course, the foregoing is mostly theoretical, and O'Keefe and Nadel had to provide evidence that the hippocampus did, in fact, function as a cognitive map. Chapters 3 and 4 of the book deal with the anatomy of the hippocampus and the physiological evidence for its role in spatial navigation, respectively, essentially providing an account of how hippocampal place cells support cognitive maps and the evidence for this. At this point, however, I want to remain at the conceptual level and discuss the idea of hippocampal neurons *representing* space. The notion of representation is omnipresent in O'Keefe and Nadel's discussion of the evidence in support of their theory. Moreover, the emphasis put on the role of map-like representations in supporting complex spatial behavior seems to suggest that O'Keefe and Nadel are not speaking metaphorically and that there is a substantive commitment on their part to the idea of hippocampal place cells doing representational work. Note that a more detailed discussion of the neural mechanisms that underlie cognitive mapping will be conducted in section 4.2 of this chapter.

4.1.2 Neural representations

O'Keefe and Nadel repeatedly bring up the idea that complex spatial behavior requires internal map-like representations of space, in part because the idea was already present in the literature on homing and migration in birds (e.g. in Griffin 1955). The entire point of the first two chapters of their 1978 book was to establish 1) that any account of spatial behavior needed to start with an account of how organisms can understand absolute space and 2) that such an account necessarily had to rely on a notion of cognitive mapping, itself understood as the capacity to build internal maps of one's environment.

Of course, O'Keefe and Nadel were interested in the neural basis of such map-like representations, and although I have focused on the conceptual background of their theory, most of their 1978 book is dedicated to that task. As radical and, in some ways, unwelcome, as the recourse to internal representational states was when first introduced in the 1940's amidst the dominance of behaviorism, the notion is commonplace in modern psychology and cognitive science, as well as in neuroscience. Neuroscience, however, is a fractured discipline, with many sub-disciplines using the notion of representation in different ways and to different ends. O'Keefe and Nadel saw hippocampal place cells as representing places in their environment, with the hippocampus itself acting as an aggregate of these place representations, a map. In this context, representations are not only a useful way to conceptualize the neural basis of complex spatial behavior, they also are explanatory targets in themselves. In cognitive neurobiology, the status of representations is not nearly as clear, as the notion seems to be used in a way that isn't fully substantive, but also in a way that is more than "shorthand" (Sullivan 2010). Philosophers of science have also been keenly interested in clarifying the status of neural representations. Mandik (2003) identifies the task of explaining what it means for neurons and neural system to have representational content as one of the key issues of philosophy of neuroscience, while some neuroscientists try to clarify the potential pitfalls and methodological demands that come with representational frameworks (Davis and Poldrack 2013). On the other hand, the very notion of neural representation is the target of no small amount of criticism from philosophers. For instance, Cao (2011) asserts that while construing the brain as a signaling system possessing semantic representational properties is a very common strategy in neuroscience, it is also mistaken. Ramsey (2007) believes that representational vocabulary is not much more than shorthand at best and misleading at worst. My aim here is to get clear on the sense in which neurons can be said to represent, and how the representational view of neural activity translates into research questions in neuroscience.

What does it mean to say that neuroscientists characterize neural activity as representational? As a first pass, here is how Bechtel describes the strategy:

> “[Neuroscientists] construe [...] neurons as *representing* those features in the environment whose presence is correlated with the increased firing and attempt to understand how subsequent neural processing utilizes representations that stand in for those features of the environment in guiding behavior.” (Bechtel 2014, p. 2, emphasis added)

This is clear enough: characterizing neural activity as representational amounts to adopting a general conceptualization of what neural activity *does*. On this view, it acts as an informational vehicle, making some content about the external world available to other neural processing systems, which in turn either use the information to regulate some aspect of behavior or communicate information to other systems, which will also either use the information or send it downstream to a control or regulation system, and so on. In addition, such representations are (by assumption) taken to be veridical (Sullivan 2010, p. 877, see also Akins 1996 for a critique of the veridicality assumption). This analysis needs to be fleshed out, however.

In a discussion of neural representation[45], deCharms and Zador (2000) synthesize a large set of views on the notion. It is interesting to note that they begin the article by stating that "[the] principle (sic) function of the central nervous system is to *represent* and transform information and thereby mediate appropriate decisions and behaviors." (deCharms and Zador 2000, 613). On their view, neural representations are defined by two characteristics: content and function. The content is the information carried by the representation. The function is the effect the signal may have on cognition or behavior. Note that this makes the property of "representing" a relational one. If a neuron has no axons, it could only be said to represent something for the experimenter since "no one else is listening" (2000, 615). In this sense, the "function" of a neural representation always depends on there being a consumer for this representation[46]. This idea is central to the notion of neural representation, as it is to that of mental representation (at least on some accounts, e.g. Millikan 1989). The notion of consumer ("function") is also quite flexible. At one point deCharms and Zador seem to understand the function of neural representations in a very general way, stating that their function is to provide information on the environment and promote adaptive behavior (deCharms and Zador 2000, p. 617). They note, however, that the nervous system carries out a host of operations on information and that consequently the concept of neural representation can (and is) used in many different contexts (2000, p.618). They go on to characterize the notion of neural representation as applying to any situation in which information is coded by neurons in a way that can be measured or analyzed and where a transformation of information takes place. It is thus a very unconstrained notion.

Another characteristic of representations that cuts across disciplinary boundaries is that they are informational vehicles: the content is formatted and carried in such a way that the consumer of the representation can use it. This is true of both mental and neural representation. In the case of neural representations, a core concern is that of specifying

---

[45] deCharms and Zador use the term "neuronal representation". To avoid confusion, I adopt Bechtel's use of "neural representation" since it fits with the philosophical literature on the topic.

[46] I use "consumer" to refer to this sense of function, as in this context "function" is essentially used to capture the consumer of a representation.

what exactly the vehicles are; how neurons transmit information. A large part of deCharms and Zador's review is devoted to a discussion of various hypotheses on neural coding (2000, pp.619-631), highlighting the substantial debate over which characteristics of neural activity serve as the actual informational vehicles. Note that neurons themselves are not the informational vehicles, it is their activity, or a specific type of neural activity, that carries information.

Summing up, neural representations are informational vehicles that carry content in a way that makes it usable by a consumer (be it the organism or a lower level consumer such as a neural system downstream). As deCharms and Zador's review shows, the main issue in the scientific literature on neural representation is that of specifying *how* neurons code information. The idea that information is coded and represented by neural activity appears widely accepted, but the Devil is in the details. Moreover, the notion seems to be highly flexible, as it applies to any context in which some form of information processing occurs. Thus, one might expect to find discussions of neural representations in more than one area of neuroscience. This suggests that whatever the status of the notion of neural representation, it is at least a very common one. Moreover, it seems appropriate to speak of a *commitment* to the representational framework. Here "framework" refers to the set of terms that characterize a system. In the remainder of this section, I (cursorily) discuss Bechtel's (2014) analysis of the case of place cells, which exemplifies the important role that the commitment to representations plays in neuroscience. I analyze selected "episodes" of place cell research to illustrate Bechtel's point that construing neural activity as representational constitutes an important commitment on the part of neuroscientists.

4.2 Bechtel's decomposition and localization analysis

The notion of neural representation discussed above has three components: vehicle, content and consumer. My aim in this section is first to show how questions about each component contributed to the research agenda on place cells, driving the development of specific research problems. I then analyze the various research avenues described here from the perspective of Bechtel's analysis of the case of place cells. As I will argue in later sections, Bechtel's analysis will be found wanting with respect to its breadth. While it provides an informative analysis of the role the notion of representation played in the development of the place cells research agenda, it leaves unanswered the question of why representations became central to its development. As I show, the dynamic-iterative model, in contrast, provides a clear answer. To be painstakingly clear: Bechtel's analysis easily explains why representation is a crucial notion, but it does not explain what led to its initial adoption, and I believe that answering *that* question sheds new light on the status of neural representations. While the previous section focused on the elaboration of O'Keefe and Nadel's theory of the hippocampus as a cognitive map, this section begins with a succinct overview of the empirical work O'Keefe and others conducted over the years. The overview is not organized chronologically, but rather thematically, around the three components of neural representations (vehicle, content and consumer).

First, there needed to be an account of what place cells were doing that enabled them to represent location. Recall that the representational vehicles are not the neurons themselves, but certain features of their activity. The initial hypothesis was that firing rate carried the information. The further from the center of the place field the rat gets, the slower the firing rate (Bechtel 2014, p. 17). This was abandoned because it was found firing rate carried non-spatial information (p. 17). A subsequent attempt to use EEG in addition to single-cell recording to link place cell firing to theta rhythms[47] (6-12 Hz oscillations) fizzled when it became apparent that there was no straightforward relation between a theta rhythm and the firing of place cells (Ranck 1973; for a more complete discussion of this, see Bechtel 2014, sec. 6; Buzsáki et al. 1983)[48]. A different avenue was pursued by O'Keefe and Recce (1993). They recorded bursts from a single place cell in CA1 while the rat ran back and forth along a track and related these bursts to the theta rhythm. On a typical transit the rat would be in the place field of a given neuron for several theta oscillations. O'Keefe and Recce noticed that the neuron's spikes were spaced regularly but at a slightly higher frequency than the theta rhythm. As the rat moved through the place field spikes would outpace the theta cycle's phase such that by

---

[47] There had been sustained interest in theta rhythms since the 1930's and researchers had been trying to tie them to mental processes such as voluntary movement (see Bechtel 2014, n. 15; Buzsáki 2005).

[48] Note that Ranck referred to place cells as "complex spike cells".

the time the rat left the place field the spikes would have advanced by almost a full cycle. This was called *phase precession* by O'Keefe and Recce and they suggested that measuring how far the spikes precessed the theta rhythms would provide a more accurate representation of the animal's position than the spike alone. Rather than the firing rate of place cells, it is the temporal situation of the spike relative to the theta rhythm that provides the fine-grain representation of location.

Second, O'Keefe had to figure out how exactly place cells could represent locations – whether they represented locations themselves or relied on a certain class of stimuli as a proxy. This question pertains to which stimuli were, in fact, being represented by the action potentials of place cells or, in other words, what the content of these representations was. O'Keefe and Conway (1978) came to the conclusion that no single stimulus was being represented by place cells. Rather, they seemed to respond to clusters of local cues (two seemed to be sufficient to elicit a response), and did not respond to distal cues such as the environment beyond the maze. This kind of progress is a direct consequence of the commitment to representations, as the representational framework provides targets of explanation (what is being represented).

Finally, Bechtel (2014) remarks that spatial representations[49] only play a role in information processing if they enable an organism to navigate its environment. This means that spatial representations must be employed by cortical areas involved in motor planning – this is a question related to the consumers of spatial representations. Some work on this issue has focused on the posterior parietal cortex (PPC), as it is heavily connected to the hippocampal regions associated with spatial representations. A promising hypothesis is that the PCC in rodents is involved in converting allocentric spatial information (developed in hippocampal areas) into a format that can be used to direct movement (p.28). As it is, there remains much work to be done to understand

[49] For reasons of space I omit the discussion of the refinement of the account of spatial representations, but a quick summary would mention that in the course of determining what sort of information was used by place cells to output spatial representations, it was discovered that they rely on the representations provided by grid cells and head direction cells (Bechtel 2014, 27). This explains why no single stimulus was found to be the determining source of place cell activity — they relied on previously integrated information.

which areas employ the representations cobbled together in the hippocampus, and place cells appear to play varied roles in a variety of cognitive processes.

These three avenues of research exemplify the way in which representational frameworks direct inquiry. Because they provide a general conceptualization of what neural activity does (transport information in certain formats for use in behavior regulation), they introduce a host of explanatory and experimental targets, the most obvious and general one being the place representations themselves, but the foregoing shows that there were more specific targets depending on the context: vehicle, content and consumer. Bechtel's analysis of place cell research supports the idea that representations play a substantive role in neuroscience research, one that goes far beyond "mere" convenience.

On Bechtel's view, the commitment to the notion of representation specifically allows for the construction of mechanistic accounts in neuroscience. A mechanistic account is one that identifies a phenomenon of interest and specifies how this phenomenon is produced and maintained by specific parts of a system. This kind of explanatory account requires that the system is decomposed into its component parts and that each function that is thought to be involved in producing the phenomenon be localized to a specific part, or set of parts, of the system. Using the overview of place cell research above, this means that the discovery of place cells that reacted to specific place fields in the animal's environment amounts to a "*direct localization* of a map that provided an allocentric representation of space" (Bechtel 2014, 30). This localization claim merely identified where in the system (brain) the phenomenon of interest was produced. The system was further decomposed, as detailed above, into those parts that acted as vehicles, those parts that provided content to the vehicles, and those that received and used the information carried. This motivates Bechtel's view that neuroscientists are committed to representations and that talk of representations is not a mere gloss, but a substantive commitment. That commitment is substantive because it allows for strategies such as decomposition and localization to be put to use in constructing a mechanistic account of spatial navigation.

While this picture is certainly accurate, it leaves out an important aspect of the analysis. Representations are not only explanatory *targets*. As is evident from the discussion of the context in which O'Keefe and Dostrovsky discovered place cells in 4.2.1, the idea that complex spatial behavior relies on map-like representations of space first occurred as an explanatory hypothesis, and while it is true that representations of space became explanatory targets, they began as explanatory resources even before it was discovered that certain neurons responded preferentially to specific areas in an organism's environment. Decomposition and localization are strategies that can be applied only when there is a system to decompose and functions to localize. That is, they are heuristics applied to *empirical* problems. Recall that empirical problems, per the taxonomy outlined in chapter 3, are problems that call for a goal state in the form of an identification of the empirical substrates of the explanatory concepts brought in to make sense of a puzzling or otherwise problematic observation. What is missing from Bechtel's analysis is an account of *why* there is a commitment to the representational framework beyond the fact the framework supports the use the decomposition and localization heuristics. This may be a consequence of the fact Bechtel's analysis is couched in terms of mechanistic explanation. That is, decomposition and localization are the basic heuristic strategies employed in developing mechanistic accounts, they are the core strategies that allow for the identification of the component parts of a mechanism and of the functions these parts perform. My contention is that while it is certainly true that committing to representations supports the application of decomposition and localization, the commitment itself was not made *because* of this. The commitment to representations derives from reasons internal to the discipline, and examining those reasons reveals a different kind of problem than those for which decomposition and localization are helpful. It reveals a different class of problems for which heuristic procedures can be useful, and it also reveals how the solution to those kinds of problems provide a conceptualization of a system which can be decomposed and accounted for in mechanistic terms. The next section delves deeper in the history behind the discovery of place cells to show that the adoption of a representational framework functioned as a heuristic, one that constrained the problem of explaining complex spatial behavior such as homing and navigation in migratory birds and latent learning in mammals.

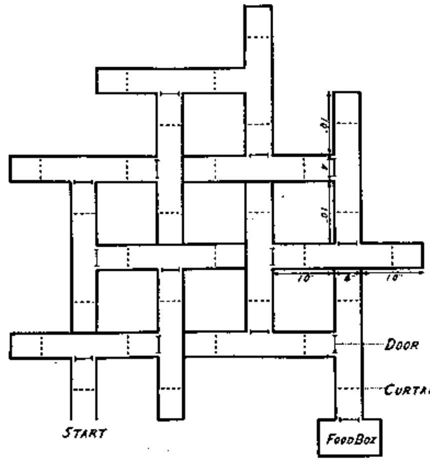
4.3 The dynamic-iterative analysis

This section is divided in two sub-sections. I first provide some historical background about the origins of the use of representational frameworks in neuroscience. This historical interlude provides contextual information that allows for the identification of the conceptual problem that stood in the way of reaching a specific explanatory goal, viz. that of explaining the phenomenon of latent learning.

4.3.1 Historical interlude

The discovery of place cells by O'Keefe and Dostrovsky (1971) lent support to the idea that spatial learning and navigation relied on internal "maps" of the environment. My aim here is to trace the origin of this idea and clarify its role in the discovery of place cells. Section 4.2 already discussed the importance of research on spatial navigation that inspired O'Keefe and Nadel to propose the idea that the hippocampus functions as a cognitive map. Moreover, at the time when O'Keefe and Dostrovsky conducted their experiments, the hippocampus was known to be involved in memory processes. For instance, Scoville and Milner's work with patient H.M. suggested that the hippocampus might be involved in the formation of new long-term memories, as their patient suffered from anterograde amnesia following a bilateral resection of the hippocampus and surrounding cortex (Scoville and Milner 1957). This background knowledge told O'Keefe and Dostrovsky *where* to look, but *what* to look for was a piece of the puzzle that was provided (at least in part) by the move away from behavioral accounts of learning.

This move has its origins in the debate over latent learning. Latent learning is a phenomenon whereby rats learn the layout of a maze without reinforcement being involved (Tolman 1948; Tolman and Honzik 1930). Tolman (1948) describes a series of experiments where hungry rats navigate a maze to get to a goal-box (see fig. 3 below). Results showed that rats made fewer mistakes after repeated trials, indicating that they were learning how to navigate the maze and knew where the food was located. More specifically, the rats' error-curves (number of mistaken paths taken) decreased progressively (Tolman and Honzik 1930).



Plan of maze
14-Unit T-Alley Maze
FIG. 1
(From M. H. Elliott, The effect of change of reward on the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1928, 4, p. 20.)

Figure 6:T-alley maze (Elliott 1928, in Tolman 1948)

The dominant approach at the time was behaviorist and held that the relevant stimuli from the maze were represented in the rat at the same time as associated responses. On this view, learning is a simple process of strengthening the neural connections that lead to desirable outcomes (finding food) and weakening those that resulted in errors. In other words, learning occurs on the basis of simple reinforcement.

Latent learning experiments modify the experimental design by delaying the inclusion of a reward in the maze. Rats go through a maze and are fed only an hour after trial completion. They exhibit no significant change in performance over the first days of the experiment, but when food is introduced in the maze their performance almost immediately improves, matching the performance of rats trained on the standard paradigm (Blodgett 1929). These results pose a problem for the behavioral account, as it seems learning can occur in the absence of reinforcement. This is, at any rate, the way Tolman interpreted these results (1948, pp.194-5).

In a critical review of modern introductory psychology textbooks, Jensen (2006) asserts that the phenomenon of latent learning is widely discussed[50] in the context of the limitations of, or progress beyond, traditional behavioral accounts of learning (Jensen 2006, p. 187). Generally, latent learning results are seen as posing a serious problem for behavioral accounts, as they do not have the resources to account for learning occurring in the absence of reinforcement or conditioning. Tolman's proposal is, explicitly, an attempt to provide the required explanatory resources to handle latent learning (Tolman 1948).

This proposal, termed "field theory", asserts that rats build up a 'map' of their environment, and that the neural processes involved in the construction of this map (and thus spatial learning) are more complex than those involved in the simple one-to-one connections between stimulus and response. Specifically, stimuli are treated and selected by a central processing mechanism, which isolates relevant stimuli and encodes them in a cognitive map of the environment. In effect, Tolman is positing internal representations to explain how latent learning could happen. Note, also, that there was some amount of resistance to the proposal. For one thing, many of the researchers who conducted the experiments on which Tolman relies did not see their results as much more than a puzzle for the then-dominant behavioral approach (for instance, Blodgett (1929) does not mention internal representations). Furthermore, as Jensen notes, some researchers provided behavioral accounts of latent learning (e.g. Guthrie 1946, 1952; Hull 1952). Note that my concern here is not with analyzing competing accounts of latent learning (and at any rate, the research agenda sparked by Tolman's proposal has certainly proved fruitful), but it is important to highlight the fact there *were* alternative accounts. Otherwise, the move of adopting the representational framework is not very interesting.

To conclude this historical interlude, I want to address a peculiarity of Tolman's 1948 proposal and draw attention to the consequences of adopting his suggestion. First, it is unclear whether Tolman proposed the existence of something like "whole organism-

---

[50] Jensen examined 48 textbooks published between 1948 and 2004, 36 of which discuss latent learning (Jensen 2006, p.187).

level" representations (which are closer to the philosophical notion of representation) or if he was using the term in the more limited sense of "neural representations"[51]. The difference between the two largely concerns the consumers and content of representations. In the case of whole-organism representations, the degree of complexity of the representation will be assumed to be greater, because it should contain all the information necessary for the organism to adequately regulate its behavior. A neural representation, on the other hand, can represent only part of the information relevant to behavior regulation and still fulfill its role in the causal economy of the brain. This distinction is important in the context of philosophical debates about the status of neural representation, as discussed in section chapter 5.

Finally, Tolman's model provided a new way to conceptualize the processes that support spatial navigation and learning, which is explicitly recognized as a strong influence on the subsequent development of place cell research (O'Keefe 2014). It also introduced a number of empirical questions (problems): how to support the hypothesis via empirical evidence, how these maps function, viz. whether they are narrow, strip-like maps (that only include information about trained-on paths) or broad-comprehensive maps (that enable flexible navigation to the goal box), as well as questions related to the neural correlates of the maps. Put differently, once the cognitive maps hypothesis is adopted, several questions become valuable research problems.

4.3.2 Analysis

What led Tolman to propose that rats build up an internal map-like representation of their environment was the perceived inability of the then-dominant behavioral approach to account for latent learning. Essentially the issue here can be understood as a discipline facing a problem of conceptualization: the available conceptual resources were not able to handle the latent learning results – at least this was Tolman's rationale for proposing the field theory model. Moreover, Jensen's (2006) analysis of introductory textbooks in

---

[51] In fact, it is unclear whether he thought such maps really existed or if he merely employed the term metaphorically. O'Keefe and Nadel seem to think the latter is more likely and explicitly adopt the substantive interpretation (John O'Keefe and Nadel 1978, p. 51).

psychology confirms that this is a widespread reading of the context in which Tolman introduced internal representations. Thus, on the dynamic-iterative model, the analysis begins with the identification of a conceptual problem:

> *Conceptual problem C1: The discipline lacks an appropriate conceptualization of brain activity that can account for latent learning.*
>
> *Goal state: Provide a conceptualization of brain activity that allows for an explanation of maze performance in rats.*

To see why the conceptualization of brain activity matters, consider the basic explanatory problem that stemmed from the observation of maze performance and latent learning:

> *Explanatory problem P1: There needs to be an explanation of maze performance in rats.*
>
> *Goal state: Provide an explanation of maze performance in rats.*

This problem specifies a clear goal state, but is otherwise significantly unconstrained. The adoption of some form of conceptual framework constitutes a necessary first step in providing a solution to problem P1. The dominant behaviorist approach rejected the use of internal cognitive states as acceptable explanatory resources. As a result, there are some conceptual elements that are not admissible as explanatory resources because of the theoretical commitments of behaviorism, which largely stemmed from epistemological concerns about unobservable internal cognitive states. Thus, one version of this explanatory problem, with the added constraint of respecting the behaviorist conceptual framework is the following:

> *Explanatory problem P2: There needs to be an explanation of maze performance in rats that does not refer to internal cognitive states.*
>
> *Goal state: Provide an explanation of maze performance in rats.*
>
> *Constraint: The explanation must not refer to internal cognitive states.*

Problem P2 is differs from P1 because of the additional constraint. Whether one adopts the behaviorist or the field theory model, the result amounts to imposing certain constraints on the search for an explanation. This is the posited function of heuristics. In the case of the behaviorist framework, the constraints dictate that explaining spatial navigation is done by identifying the stimuli that trigger the desirable response, whereas the representational framework introduces a different set of requirements (explanations have to specify the operations performed on the incoming stimuli, the resulting content of the bodies of information so produced, and how they are used). Thus, if one adopts the representational framework, that is, if one adopts a different conceptualization of brain function, then one is dealing with yet another different problem:

> *Explanatory problem P3: There needs to be an explanation of maze performance in rats that can and should refer to internal cognitive states.*
>
> *Goal state: Provide an explanation of maze performance in rats.*
>
> *Constraint: The explanation must specify the operations performed on the incoming stimuli, the resulting content of the bodies of information so produced, and how they are used.*

In both cases, solving the conceptual problem by adopting a given framework introduces a host of empirical problems. For instance, adopting the behaviorist framework introduces such problems as identifying the relevant stimuli, and the way stimulus-response couplings can be delayed in their expression (to account for latent learning). Adopting the representational framework introduces such problems as specifying which stimuli are required to allow for the construction of mental maps, and the precise nature of these maps (whether they are landmark-based, whether they are flexible, etc.). In either case, however, the adoption of a conceptual framework allowed for the formulation of a more constrained version of explanatory problem P1, and led to the identification of specific empirical problems that, as seen in section 4.2, made up the bulk of the place cell research agenda.

Insofar as it imposed new constraints on the original explanatory problem, the commitment to representations fulfills the posited role of heuristics in science. Indeed, even though the representational framework represents an opening of explanatory possibilities *in the context of a field dominated by behavioral accounts*, it also closes off other research avenues. Alternatives to the representational frameworks are those based on dynamical systems theory (e.g. Chemero 2000). Moreover, the notion of a cognitive map of the environment is, after all, quite vague, and does not by itself provide a complete explanation of spatial navigation. In fact, research on place cells uncovered a complex organization of cells that contribute in different ways to the production of spatial representations. This is because, as discussed in chapter 3, heuristics create new problems. Applying a heuristic to a problem transforms it into a new problem in virtue of heuristics imposing more constraints on an existing problem, which by definition results in a new, hopefully related problem. However, this transformation may create other problems downstream. As an example, think of the reductive localization heuristic, whereby a relational property involving organism and environment is treated as a monadic property of the organism. This might well make the problem of understanding fitness more tractable, but at the cost of creating the problem of reconciling the resulting account of fitness with the fact organisms evolve in an environment. Such problems may become part of the research agenda (as was the case with the problem of identifying the proper vehicles of the spatial information involved in constructing spatial maps), or they may be left out altogether. An evolutionary biologist might consider the problem of reconciling notions of fitness with environmental considerations to be the job of the ecologist, for instance. Similarly, the commitment to representations introduces productive research problems (as evidenced by the discussion in this chapter, in particular section 4.2), but also more difficult problems that may fall outside the scope of inquiry (e.g. reconciling neural and mental representations, see section chapter 5)[52]. Let us look at some examples.

First, the representational framework introduced questions about how neurons encoded information, that is, what feature of their activity acted as a representational vehicle. Such a question can be reformulated as an empirical problem:

> *Empirical problem E1: The empirical correlates of the concepts involved in the representational framework need to be identified.*

---

[52] The philosophical notion of mental representation is stronger than the neuroscientists' notion of neural representation. The mismatch between these notions is the basis of Ramsey's (2007) argument that neural representations are not true representations and consequently cannot be more than an interpretive gloss. This is an interesting question but different than the one I tackle here.

> *Goal state: Identify the feature of neuronal activity that carries information.*

Much like problem P1 above, this problem is unconstrained. However, because of the disciplinary context, and especially the pragmatic aspects of single-cell recording, it is possible to identify at least one implicit constraint, one that motivated the formulation of the initial hypothesis that firing rate carried the relevant information, and thus initial experiments directed at identifying the vehicles of spatial representations focused on firing rate. In other words, the problem was constrained by decomposing the system into parts (in this case, neurons or populations of neurons) and localizing the function of coding information in the firing rate of those neurons:

> *Empirical problem E2: The empirical substrate of the concepts involved in the representational framework need to be identified and are likely related to firing rate.*
>
> *Goal state: Identify the feature of neuronal activity that carries information.*
>
> *Constraint: The firing rate of neurons is the main experimental target, viz. experiments should focus on measuring firing rate in relation to behavioral observations.*

These empirical problems were *generated* by the previous application of a heuristic: it is only because the representational framework had been adopted in the first place that such a problem was part of the research agenda.

Conversely, some problems that are generated by the application of a heuristic either go unnoticed or simply fall outside the domain of enquiry entirely. An example of the latter is the mismatch between philosophical notions of representation and the notion of neural representation. Some challenges to the use of representational vocabulary in neuroscience (Ramsey 2007; Cao 2012; Haselager, Groot, and Rappard 2003) point out that it is a mistake to conceptualize brain activity as representational because the only sense in which neurons can be said to represent is in a minimal, explanatorily irrelevant sense. Ramsey argues that the use of representational vocabulary in neuroscience is never directly descriptive, and is a “mere” gloss. I discuss this debate at length in chapter 5, but for now it is important to note that the discussion conducted in this chapter shows that neuroscientists are indeed committed to their representational framework, and that this framework fulfills a much more substantive role than Ramsey suggests. Nevertheless, the conceptual mismatch between the two kinds of notions is a legitimate conceptual problem, albeit one that falls outside the scope of inquiry. I provide an example of the other kind of problem (the unnoticed problem) in the paragraphs below.

Committing to a representational framework, useful as it may be, may fail in characteristic and systematic ways. Recall that one of the main advantages of the notion of heuristic is that because of their lack of flexibility it is possible to predict, or at least recognize, their potential failures. The discussion in 4.2 focused on showing how the basic concept of representation, which includes the notions of vehicle, content and consumer, provided direction for research on place cells. On the flipside, one can safely assume that a research agenda based on a representational framework will exhibit certain kinds of failures related to these three notions. The tale of place cells is replete with misidentification of the empirical correlates of these notions. For instance, the initial hypothesis regarding the vehicle of spatial representations was that firing rate carried spatial information, which turned out to be wrong. Similar instances can be found with regards to content and consumers (see section 4.2). Also noteworthy are more general (and arguably more serious) errors, which relate to the use of representational frameworks. It might be the case that the observer attributes a certain content or role to a neural event that is not itself used by the system under study. In other words, the only consumer of a representation might be the experimentalist. A potential issue is whether the phase precession is really the vehicle used to transmit spatial information *for the system* or if it is only "consumed" in this way by the researchers interpreting the data[53].

A similar issue comes up when considering the matter of context of use. Just as the representativeness heuristic (Kahneman et al. 1982, pp. 153–163, see also chapter 3), the representational framework might fail when applied to different contexts. Wright

---

[53] I owe this point to Wright's (forthcoming) discussion of MVPA. Note that this is a difficult question that would require its own paper. I nonetheless bring it up, as it shows how conceiving of the representational framework as a heuristic liable to fail in certain ways can indicate interesting avenues of research.

(forthcoming) discusses the use of multi-voxel pattern analysis (MVPA) in attempts to identify the content of representations coded in the neural signal (Haxby et al. 2001). They showed that a machine learning classifier could predict what subjects were viewing with 90% accuracy and from there inferred that the technique could specify what the content of the representation was. Wright notes, however, that there is an "inferential distance" that is left unexamined, as there is no guarantee that the system studied has access to the same information as the classifier. In such a case, the error stems from having the wrong consumer for the representation. That is, there *is* a representation, but the representation is not used by the system at all, only by the classifier and, by extension, by the observer. This is the difference between a representation *gleaned from* the system and a representation *used by* the system. Moreover, it might well be the case that the representational framework is entirely the wrong way to conceptualize neural activity and that proponents of dynamical systems approaches such as Chemero are correct. I make no claim that the representational approach is the right one, only that 1) it has proved to be a productive one and 2) that this is because it functions as a heuristic. Whether it is right or wrong is another question altogether (what's more, heuristics do not offer guaranteed success).

Note that these errors are part of the equation: the initial conceptual framework which generates the empirical problems that lead to the gathering of relevant data and observations is modified per the new information gathered, and then this new conceptual framework generates a new batch of empirical problems, and the process starts over again. This is the iterative part of the model of scientific practice. As more information is gathered, more constraints are added to the initial research problem. The dynamic-iterative model allows for a complete description of the process, and could easily be adjusted to focus on one particular episode in the iterative process, or, as I have done here, used to provide a general description of a research agenda.

4.4 Conclusion

At this point, I believe I have shown that the dynamic-iterative model is in fact applicable in practice. The taxonomy of scientific problems outlined in chapter 3 allowed me to account for the initial adoption of the representational framework in neuroscience in a way that explained why research on place cells proceeded in the way described in Bechtel's analysis of the case (2014).

The novel claim I made in this chapter is that the commitment to representations is a heuristic, but one that differs from those discussed by Bechtel (2008, see also Bechtel and Richardson 2010). First, the commitment functioned as a source of new constraints on an explanatory problem. Second, the commitment showed all of the other features of heuristics identified in chapter 3. I believe that these are reasonable grounds for supporting the claim that the commitment to representations is a heuristic. Interestingly, this commitment, while exhibiting the characteristic properties of heuristics, also exhibits key differences with the heuristics of decomposition and localization. Specifically, decomposition and localization are heuristics that can only be applied once there is a conceptualization of a system in place. The adoption of a conceptual framework is a prerequisite for the application of heuristics such as decomposition and localization, and it is in this sense that the dynamic-iterative analysis is both compatible with, and a more complete analysis than, Bechtel's mechanistic analysis of the case of place cells. Although, as I discuss in chapter 5, the connection to mechanistic explanation is a contingent one, and there is much to gain from removing the analysis of heuristics in science from the context of mechanistic explanation.

Chapter 5

5 Evaluating the model

In the previous chapter I showed how the dynamic-iterative model can account for the use of representational frameworks in neuroscience in ways that current analyses cannot. This chapter will focus on taking stock of what has been learned from the application of the model to the case study in terms of the model's strengths and weaknesses and how it compares to similar approaches in philosophy of science.

The chapter is divided in four sections. In section 5.1 I bring up possible objections to various aspects of the model. Responding to these objections allows me to make some conceptual revisions which will be important in the evaluation of the model. Section 5.2 deals with the evaluation of the model with respect to the six adequacy conditions I laid out in chapter 1. In section 5.3 I discuss the model's strengths and weaknesses. Finally, in section 5.4 I compare the dynamic-iterative model to Bechtel and Richardson's (2010, expanded on in Bechtel 2008; 2014) problem-solving approach and show that while my model and theirs share many features and motivations, the dynamic-iterative model presents substantial advantages over theirs, and that those advantages are a direct consequence of the model's characteristic strengths.

5.1 Dealing with objections

The dynamic-iterative model is based on the thesis that scientific practice can be accurately described as composed of organized activities, which themselves are best understood as instances of problem-solving. This thesis also serves to specify what a workable description of scientific practice needs to include: an account of what problems are (the constraint inclusion model) and an account of how problems are solved (the heuristic account). Once these two conceptual components are secured, the next step is to figure out how the various problem-solving activities are organized such that together they allow for the development of research agendas, or lead to progress, or to the development of new problems which in turn form the core of subsequent scientific work. Moreover, there needs to be a conceptual framework that makes the construal of practice as problem-solving *applicable* to actual cases, something that allows philosophers of science to identify relevant problems and situate them in the network of problems associated with a specific case or episode. This is the role of the problem taxonomy outlined in chapter 3, which makes it, in a way, the most important part of the dynamic-iterative model, because it is the component that allows the model to 'get started'. Moreover, the notion of a 'goal' is crucial to the model. Therefore, I will discuss potential objections to the model that focus on these key aspects of it. These objections could have been formulated before applying the model to the case study, but discussing them *after* seeing how the model works when applied to an actual case allows my responses to also function as revisions to the model, in addition to making the objections much clearer since I am able to point to specific examples taken from chapter 4 to both articulate them and deal with them.

5.1.1 The model leaves out explanatory problems

The first step in the analysis of the case of place cells (specifically, section 4.3.2) was the identification of a conceptual problem (C1), which was defined by the following goal state: *provide a conceptualization of brain activity that allows for an explanation of maze performance in rats.* It was necessary for Tolman and others to solve this conceptual problem, because, first, there was a pre-existing *explanatory* problem, which was that of providing an explanation for maze performance in rats, and, second, the disciplinary context was such that some researchers felt that the dominant behaviorist framework could not explain the observations of rat behavior that suggested learning in the absence of direct reinforcement could occur, at least not straightforwardly.

A potential objection to my analysis is that it seems there is a category of problems, viz. explanatory problems, that was not included in the taxonomy in chapter 3 despite playing a crucial role in the development of research on spatial learning and navigation. If conceptual problems can only be formulated in the context of trying to solve an explanatory problem, my model appears to suffer from a serious weakness: it leaves out explanatory problems despite needing that category of problems to make sense of scientific practice.

The reason explanatory problems were not part of the problem taxonomy outlined in chapter 3 is that such problems are best understood as a statement of research goals as opposed to a statement of a problem that plays a role in the iterative process described by my model. An explanatory problem, on this view, is a specific formulation of the goal pursued by scientists in the context delineated by a case study. In chapter 2, and particularly in the discussion of Nickles' constraint-inclusion model, I emphasized the importance of descriptive flexibility with respect to the way these goals are characterized. Overly specific statement of the goals of science had been shown to lead to issues with descriptive accuracy in the case of Laudan's model (1978). The claim that problem-solving is, itself, the goal of science, makes Laudan's model of progress unable to explain *why* scientists want to solve problems in the first place. Similarly, the traditional approaches criticized by proponents of the Practice Turn overemphasized the motivational force of the large-scale, idealized goals of science. To 'explain the natural world', for instance, is certainly part of the goals of science in some way or another, but too far removed from day-to-day practice to be relevant in our analyses of scientific practice as it happens (see 2.1 and Rouse 2002, 104). Therefore, in most cases, an explanatory problem will be construed as a general goal.

5.1.2 The model cannot account for the motivational aspects of practice

Explanatory problems are still problems, however, and this suggests another potential objection to my model: if the goals pursued by scientists can be couched as problems, then am I not facing the same issue as Laudan faced, in that the most general goal of science is the solving of problems? In effect, this amounts to the criticism that the dynamic-iterative model reduces the goal of science in general to the solving of problems, which leads to the same issues identified with Laudan's model. This objection is based on a misunderstanding of the difference between characterizing problem-solving as the aim of science in general and characterizing science as being made up of problem-solving activities. The explanatory problem at the core of Tolman's research is only a relevant problem *because there was the need for an explanation* of the phenomenon of spatial learning in rats. The real goal is that of providing an explanation, *not* of solving a problem. The fact that such goals can be stated *as* problems does not mean that they have no epistemic significance beyond 'solving a problem', and the motivations that drive researchers to look for these explanations go beyond the motivation to simply solve a problem. As an example, consider any one of the problems discussed in the previous chapter. The problem of identifying how place cells encode spatial information (that is, the empirical problem of identifying the empirical substrate for the conceptual component of representational vehicles) is only a relevant problem because within the disciplinary context the identification of which features of neural activity acted as representational vehicles was an important step towards the goal of *explaining* how place cells functioned within the causal economy of the brain. The *motivation* for solving the problem is what matters, the goal pursued, and the problem itself is an obstacle that crops up in the process of achieving the goal state. Thus, my model does not face the issue of not being able to account for the motivational aspect. The response to this objection is that my analysis characterizes research goals in terms of problems, as opposed to asserting that solving problems is the goal of research.

5.1.3 The distinction between goals and problems is unclear

A further potential objection, which is likely to come up in reaction to the above, is that the distinction between "goals" and "problems" is at best unclear. If large-scale explanatory goals can be formulated in terms of a problem with a goal state and constraints, the worry is that there is no principled way to discriminate between those problems that count as goals and those that do not. My answer to this worry is that the distinction between goals and problems is contextual and depends on the grain of analysis, or "zoom" adopted by the researcher applying the model. Something that is better understood as a goal at one strata may become a problem at a more general level of analysis. For instance, one could take any of the specific empirical problems discussed in the previous chapter and contextualize them as goals. The identification of those features of neural activity which encode information about space was, in the context of analyzing the larger case of place cells, an empirical problem that was derived from the adoption of the representational framework. However, one might want to analyze the strategies that enabled researchers to identify possible features of neural activity that acted as representational vehicles, or even the strategies they used to generate hypotheses about those features, and in such a case the goal towards which the problem-solving activities are directed is that of identifying the relevant features of neural activity. The distinction between goals and problems is thus a contextual and epistemic one: what counts as a goal will depend on the epistemic aims of the researcher rather than on any intrinsic feature of a given problem[54].

This contextual definition of the distinction between goals and problems has two further applications within the context of philosophy of science. First, it supports the aim of developing a general descriptive framework that preserves the insights of the Practice Turn while avoiding the risk of excessive specificity. This is because it is possible to define a problem as a goal at any level that is deemed relevant while maintaining the general structure of the model. Second, the distinction enables an understanding of the relationships between layers of scientific practice. As an illustration of what I mean, think of Bechtel's analysis of the role of representation in neuroscience (2014) and how this analysis relates to my analysis of the origin of the commitment to the representational framework. Bechtel argues that the representational framework supports the strategies of decomposition and localization by providing clear explanatory targets (Bechtel 2014, 30). In other words, the relevant goal on Bechtel's analysis is that of explaining the mechanisms responsible for the production and use of representations of space. In contrast, my analysis looks at the origin of the commitment, which it explains by pointing out that the adoption of the representational framework was a consequence of pursuing a different, though related, explanatory goal (that of explaining the surprising observations of what became known as latent learning). What looks like an unexamined assumption on Bechtel's analysis (the idea that brains use representations) becomes an attempt at solving a conceptual problem (that of the dominant conceptual framework not being able to account for the observations), which *then* leads to the formulation of the empirical problems that made up the bulk of the research into place cells and their function. Similarly, one could analyze the specific experimental practices that allowed for the key discoveries made in the case of place cells by focusing on a specific practical problem

---

[54] The related objection that this implies the possibility that there is always a finer grain of analysis is one I do not wish to discuss at length here. My response is simply that there are disciplinary aims and conventions that impose a soft limit on how detailed of an analysis will remain relevant. Moreover, the fact the model can theoretically support very detailed analyses of case studies and is not locked at one level is an advantage.

and contextualizing it as a goal rather than as a problem. The contextual characterization of goals and problems thus makes it possible for the dynamic-iterative model to do more than provide a useful description of scientific practice. In the context of philosophical analysis of practice, it also allows for an understanding of how the different layers of scientific practice interact.

To make this idea clearer, consider the model diagram provided in chapter 3:

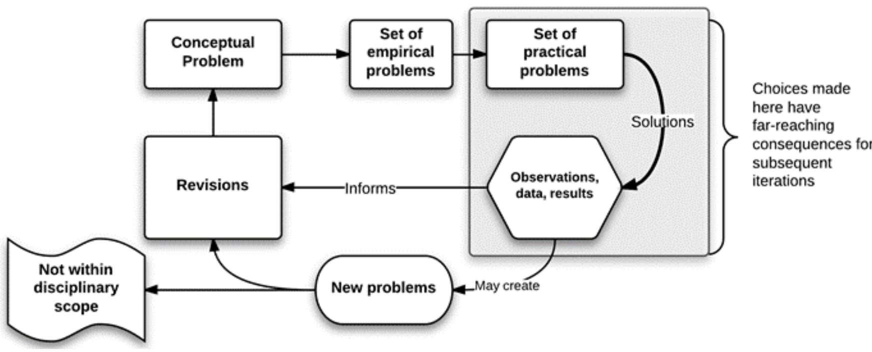


Figure 7: Dynamic-Iterative model of scientific practice

The analysis in chapter 4 began with the statement of a conceptual problem as a starting point. Following the discussion above, in the case of Bechtel's analysis (2014), the starting point was an empirical problem. Another possibility would be to begin with a practical problem if, say, the researcher's focus was on experimental practice. Each of these types of problems may be construed as goals depending on the context of analysis (i.e. what the researcher's aims are).

To make a long story short, all goals are problems but not all problems are goals. Which problems are goals will depend on the aims of researchers applying the model, what their interests are, and what their goals are in analyzing this and that case.

5.1.4 The model can only account for explanatory goals

In chapter 4, the case study focused on how the development of place cell research could be analyzed in terms of problem-solving and heuristics, and the historical analysis of Tolman's proposal emphasized the role of conceptual problems for understanding how the representational framework, which supports the use of strategies such as decomposition and localization (see, e.g., Bechtel 2014). However, this conceptual problem was relevant because it stood in the way of a specific explanatory goal, and my model is in fact vulnerable to the objection that so far, the focus has been almost exclusively on explanatory problems and goals. While I am not able to provide as detailed a demonstration as I did in chapter 4, I believe I can show that the model's descriptive scope goes beyond explanatory goals.

First, the model requires nothing more than the possibility of formulating a research agenda in terms of goal states and constraints. In what follows I discuss another case to which the model could be applied that does not involve explanatory aims, at least not straightforwardly (it could be said to tend towards explanatory goals, but only insofar as there is a case to be made for the idea that in general science tends towards the goal of explaining phenomena).

The case is that of connectomics, which is the production and study of connectomes — maps of the synaptic connections within an organism's nervous system. Connectomics has been the subject of debate within the scientific community (Morgan and Lichtman 2013), some of which reached the popular press (Costandi 2012; Cross 2016; Jabr 2012), largely because of the large sums awarded in the form of research grants to various connectomics projects in spite of serious doubts concerning the relevance of anatomical maps to the understanding of how the brain functions (Morgan and Lichtman 2013, 494). My aim is not to come down on one side or the other. Rather, I want to examine the debate about connectomics because illustrates the fact that it is a large-scale scientific enterprise with no direct explanatory goals: its goals are predictive as well as focused on the mapping or the brain. Moreover, critics of connectomics are keen to point out myriad problems the field faces, which in effect provides a ready-made list of research problems that could be analyzed using the dynamic-iterative model. Following the Human Connectome Project NIH grants awarded in 2011, several initiatives were undertaken to map out macroscopic human brain connectivity (see Craddock et al. 2013, 535, table 2 for a list of initiatives).

An overview of a 2013 review paper concerning the progress of the Wu-Minn project supports my characterization of connectomics as a non-explanatory enterprise. Although the goal is that mapping the human connectome will ultimately serve as an invaluable explanatory or predictive resource, the literature on the Human Connectome Project (HCP) makes clear that the main goal of the project is to produce something like an atlas of the connective structure of the brain (Van Essen et al. 2013; Kelly et al. 2012). In this case, a formulation of the goal of the HCP in terms of a problem would not include an explanation as the goal state pursued by researchers. The most common criticism of connectomics and other large-scale imaging efforts is that it is similar to "prying the lid off a computer to see how everything is connected" (Morgan and Lichtman 2013, 498), which is unlikely to tell you much of anything about how the computer does, in fact, work. Proponents of connectomics reply that while this is true, looking at the computer's components is still a good first step. I thus take it as fairly well-established that the overarching goal of the field is to develop a data set, which may well end up being invaluable for clinical and research applications, but in the end, is essentially "just" a map. In the pursuit of that goal, a number of large-scale conceptual problems are likely to be encountered. For instance, critics of connectomics often emphasize the difference between circuit structure and circuit function (Morgan and Lichtman 2013, 494). Since the functioning of the nervous system heavily relies on extrinsic input being processed and driving the generation of behavior, there is doubt that mapping its structure would shed much light on how it actually functions. Another issue is that the same structure can perform different functions depending on where the input is coming from, or other such factors (Morgan and Lichtman 2013, 496). Moreover, the project of mapping the human connectome forced neuroscientists to think about the relationship between different scales of description and other conceptual problems (Sporns, Tononi, and Kötter 2005), in addition to having to figure out what the best way to actually record brain activity was, given that the goal was to produce a connectivity map as opposed to identifying the neural correlates of a specific cognitive capacity.

This last problem is closer to an empirical problem than it is to a conceptual one. The goal state pursued is that of identifying what the empirical work should focus on, what, exactly, should be understood as the relevant kind of activity that would illuminate the brain's connective structure. The solution was to focus on resting state activity in the brain, activity that is intrinsic to the brain itself and that is task-independent (Kelly et al.

2012). From this solution to the empirical problem of determining the relevant type of activity to be measures came a host of practical problems, such as concerns about sample size, concerns about the possibility of controlling for movement and other perturbating factors, concerns with the plurality of data processing strategies and the lack of a standardized method, and so on (Kelly et al. 2012, 182-184). The connectome case at least appears as if it could be handled by the dynamic-iterative model.

5.1.5 The model imposes a rigid sequence on scientific practice

Finally, it may be argued that my model imposes a rigid sequence on scientific practice in that the problem taxonomy is presented as hierarchical, which conceptual problems at the top. This is a valid concern, but one that is also easy to dismiss, especially given what has been said so far in this section. By and large the most important aspect is that there is a kind of sequence through which scientific practice moves, and that each step can be characterized as an instance of a problem-solving activity. From there, the precise ordering of problems is not restricted to the conceptual-empirical-practical sequence outlined in previous chapters. The core idea of the dynamic-iterative model is that such a sequence can be identified. That is, all the model needs to 'get going' is the identification of a goal and of the problems that stand between that goal and those pursuing it. It may be that the main problem that derives from that goal turns out to be empirical rather than conceptual, but the model provides the necessary tools for identifying the kind of problem and the strategies employed to solve it. From there the model proceeds in the same way as before. Essentially, the sequencing of the process represented in figure 5.1 is flexible in the sense that what really matters is that there is a sequence. Moreover, it could be the case that solving a conceptual problem introduces practical problems. For instance, a conceptual framework could be adopted that includes restriction on what counts as acceptable experimental results, effectively 'skipping' the empirical problem stage. Note however, that this would still create a new set of empirical problems, since even if the empirical problems are the 'same' between two iterations, the second iteration of problems will have different constraints.

5.1.6 Summary

The aim of this section was to deal with five potential objections to the model. Dealing with each of those objections allowed me to clarify some key aspects of it, and those clarifications will be invaluable when evaluating the model in later sections. These five conceptual clarifications are laid out in the table below.

Table 7: Conceptual clarifications to the model

Explanatory problems are goals	Explanatory problems have a different status than conceptual, empirical and practical problems in that they are often playing the role of an overarching goal.
Status of problem-solving	The fact goals can be understood as problems does not mean the model asserts that the goal of science is problem-solving.
Contextual distinction between goals and problems	The distinction between goals and problems is contextual and depends on the aim of the philosophical analysis of a given case.
Descriptive limits of the model	The model is not limited to accounting for cases in which the goals are explanatory.
Sequencing of the iterative process	The sequence is not rigid. What matters is that there can be an identifiable sequence.

Before moving on, I want to insist on how important one of these clarifications is. The idea that all goals are problems is crucial to what will end up being the model's main strength: its flexibility. Being able to shift the model's focus between different layers of scientific practice is an important feature. This contextual aspect also applies to the class of constraints deemed relevant by users of the model. I now turn to the evaluation of the model with respect to the adequacy conditions laid out in chapter 2.

5.2 Meeting the adequacy conditions

My task in this section is to evaluate the dynamic-iterative framework with respect to the six conditions for an appropriate descriptive model of practice outlined in chapter 2. First, here are the six conditions I want to evaluate the model against:

1. Descriptive adequacy with respect to pragmatic use of theories, hypotheses, etc. This condition is derived from shift to considering theories as tools with pragmatic utility as opposed to the finished product of scientific inquiry (Woody 2014, 123-127).

2. Allow for empirical investigation through flexibility of application. This condition derived from shift to empirical from a priori methods (Woody 2014, 123-127).

3. Take into consideration the psychological factors at play in scientific practice. This condition is derived from the shift to naturalistic conceptions of scientists and their capacities (Woody 2014, 123-127).

4. Take into consideration the social dimension of science and be sensitive to the importance of contextual factors both within and without the scientific sphere. Condition derived from the shift to social epistemology (Woody 2014, 123-127).

5. The model must be applicable in practice, which is to say that it has to be actually applicable to case studies not only in terms of capturing relevant aspects of scientific practice, but in terms of not making unreasonable demands from philosophers of science with respect to how much work needs to be done for the model to be applied. This condition is derived from the overarching goal of the dissertation to develop a model that can facilitate direct engagement with non-academic contexts.

6. The model must at least provide the resources for the elaboration of normative work. This condition is derived from the anticipated objection to practice-oriented approaches that they reduce philosophy of science to “mere” description. Showing that a descriptively accurate framework can ground normative work while eschewing the top-down normativity associated with traditional approaches is, I believe, a relevant response to make to the objection.

I have already discussed how the dynamic-iterative model meets the first four conditions in chapter 3, but considering the revisions made to the model in section 5.1, I believe it is appropriate to briefly revisit them.

The dynamic-iterative model handles the first condition in virtue of the role theories play in the description of scientific practice as problem-solving. While theories can be considered the "products" of science, they also figure prominently in the analysis of problem-solving episodes, either as constraints that are used to narrow down the search for a solution (as in the case of Tolman's field theory being used as a way to delineate a new conceptualization of how brains function), or as core goals because of their role in scientific practice, and the model's focus is on how practice happens and what tools are used while solving the problems that constitute research agendas.

The second condition is also handled easily by the model. The only requirement for the model to be applicable to a given case is that there is an identifiable goal state. Moreover, the model is not restricted to a particular level of scientific practice. The contextual distinction between goals and problems discussed in the previous section makes it possible to analyze practice at different levels, while maintaining the possibility of analyzing the relationship between different layers of practice. For instance, the overall goal in the case study discussed in chapter 4 was the explanation of spatial learning, but as was discussed, the analysis could have focused on any of the problems that made up that research agenda, contextualizing that problem as the overall goal and analyzing how *that* goal was attained. Thus, the model meets the flexibility condition.

As far as the third and fourth conditions go, the model's reliance on the notion of heuristic, as well as its emphasis on contextual factors ensure that the model complies with both the need to recognize the importance of the psychological constraints with which scientists work and the social dimension of knowledge. Thus, the model meets the four conditions derived from the methodological tenets of the Practice Turn.

The previous chapter showed that the model can, in fact, be applied fruitfully to an actual case. Moreover, I have gestured towards another case the model could apply to (see 5.1). The most important feature of the model with respect to its in-practice applicability is that each step of its application is straightforward and clear: identify the overall goal, identify the problems that crop up in the pursuit of that goal, and how they were solved. Of course, the application of the model to other cases may reveal difficulties, but within the limits of the dissertation, I believe the model's applicability is robust enough to at least warrant further 'trials'.

To demonstrate that the model meets the last condition, I wish to address the concern that the model is merely a set of tools that can be applied to scientific inquiry as a way to make it intelligible. The reason this might be a problem is that 'mere' description may be deemed insufficient in terms of philosophical relevance. This is a long-standing concern in philosophy of science. Some "new mechanists" (Craver 2007; Kaplan and Craver 2011) consider it necessary to provide clear demarcation criteria for good explanations, for instance, and are quick to point out that the mechanistic view of explanation they develop can provide such criteria. In other areas, philosophers are often wary of simply "deferring to science" in matters of normativity (see for instance Magnus' 2003 discussion of underdetermination in physics). That is, "leaving the normative to the scientists" amounts to discounting the possibility of substantive normative contributions from philosophy, which makes philosophers of science not much more than observers who lack an active role in science. The fact discussions in philosophy of science very rarely directly impact scientific practice notwithstanding (see Douglas 2010 as well as the introduction to this dissertation), this is a valid concern. More so since I have repeatedly characterized the dynamic-iterative model as a *descriptive* framework. This being said, I believe the model has important normative implications that go beyond the traditional construal of normativity in the sense of providing ways to evaluate scientific research. The model also provides meta-theoretical recommendations for philosophy of science itself, as well as ways to actively engage scientists.

First, the model provides tools to reconstruct the historical development of a research agenda. In the case of place cells, had the research agenda not been fruitful, treating its background adoption of the representational framework as a heuristic would point to possible causes for this "failure". When such stumbling blocks come up, their origins may be buried deep in the history of a research agenda and this is where the historical character of the problem-solving model comes into play. It allows the reconstruction of an agenda's development in such a way as to identify the problems that scientists faced, how they attempted to solve those problems, and which of these solutions ended up being foundational building blocks in later work, such as with the adoption of the representational framework. In a sense, this gets at something like Wimsatt's (2007) notion of generative entrenchment, which is the idea of certain structures being not only deeply nestled in a given system but also being "core" structures, viz. structures on which a lot of other elements of the system depend. Thus, the model not only enables one to understand which questions were posed and why, but also how their answers shaped the subsequent development of research. At the same time, this makes it possible to pinpoint errors and provide substantive normative input.

This is because the notion of heuristics provides the model with a built-in source for normative recommendations. One of the main features of heuristics is that they are liable to systematic errors. They are useful precisely because of their lack of flexibility: they tend to always exclude the same type of information, or close off an entire class of possible explanations (e.g. as was the case with the adoption of the behaviorist framework, which automatically closed off any explanation of spatial learning in terms of representations). This means that it is possible to identify the kind of error that a heuristic is liable to lead to. For instance, Wimsatt (2006, 2007) points out that a common heuristic for characterizing properties of systems is to keep the environment static, which makes it easier to model the behavior of the system. However, this strategy may lead to errors in cases where important properties of the system are *relational*, such as in the case of the property of fitness: an organism is never fit in a void, it is only fit in an environment, which is not static. Thus, the fact heuristics are biased provides a way to formulate normative recommendations: if it is possible to identify the ways certain heuristics might fail, then there are specific areas to which philosophers of science can draw attention when assessing a research agenda, in addition to being able to identify possible errors in the historical record[55]. The model is thus not merely descriptive: it contains the basis for a normative framework. Moreover, whatever a particular bias is a liability will depend on

[55] This same property of heuristics grounds Kenyon and Beaulac's discussion of strategies to suppress the effects of implicit bias (2014). Implicit bias functions in much the same way as heuristics: it derives from similar cognitive mechanisms. Given this, it is possible to identify the ways in which it manifests and devise, for instance, institutional systems that minimize the effect of implicit bias.

contextual elements, such as the overall goals of a scientific community, and as such, the normative potential of the model is not tied to a pre-existing view of what science should be or how it should be conducted. Rather, the model includes tools for developing normative claims that are relevant to scientists in a given context. I discuss this idea more in the next paragraphs.

The second sense of normativity at play here might be more easily grasped from the way it relates to the first sense. The normative recommendations derived from the historical reconstruction would be, in an important way, informed by familiarity with scientific practice itself. Thus, philosophers of science do not devise abstract evaluative criteria that are then applied to science in a decontextualized manner. Rather, norms of explanation or experimentation (for example) are built up from an observation of scientific practice itself. It should be noted that this is not a new idea — Craver does try to break with the "normativity from without" tradition of 20^{th} century philosophy of science (2010). Whether these attempts succeed is a question I will not try to answer here. As Chirimuuta (2013) shows, however, the application of the "mechanistic validity" criterion (the degree to which a model resembles the target system) that Craver (2010) proposes faces the problem that it essentially undermines most research in systems neuroscience, while the organizational validity criterion, that is, whether models preserve the organizational principles of the target system, Chirimuuta proposes follows this method of building evaluative criteria "from within". So, the second normative aspect of the model comes in the form of an appeal to prudence when devising normative schemes that are aimed at scientists. They should begin with a careful examination of scientific practice and, above all, the identification of the goals pursued by scientists. Of course, this does not mean that the goals themselves are exempt from criticism, but even criticizing these goals requires an initial identification of the goals actually pursued, not goals assumed to be pursued.

I believe the model presents enough normative potential to avoid the charge that it amounts "deferring to scientists". The model's aspiration to descriptive accuracy is not to be construed as an abandonment of normativity, but rather as in service of the relevant kind of normativity. This normative potential, while already present in Bechtel and Richardson's model, is bolstered by the addition of the categories of conceptual problems and heuristics. With this, I believe I have shown that the dynamic-iterative model fulfills all the conditions for a descriptive model of scientific practice.

5.3 Strengths and Weaknesses

At this point, I believe it is important to assess the model's strengths and weaknesses. I begin with a discussion of the general advantages of the model as well as accompanying drawbacks, then discuss how the model offers massive opportunities for comprehensive analysis of scientific practice at different levels, something that the models discussed in section 5.4 cannot offer, at least not as readily.

The main advantage of the model is its flexibility, which is a direct consequence of its relative simplicity and conceptual economy. Because it begins with the basic assumption that scientific practice can be broken down into separate problem-solving activities, the result is that one can apply the model to any instance of practice as long as it is possible to identify the goal pursued in the case analyzed and there is sufficient information to tease out the specific problems that stand between scientists and said goal. From there the "deployment" of the model follows a fairly well-defined sequence of identifying the relevant problems, the heuristics used to transform these problems into tractable and fruitful research problems, and the resulting revisions to the initial problem which constitutes the goal of the, and so on. The conceptual apparatus required to accomplish this is fairly limited, and the interactions between the core concepts have been discussed at length in this chapter and others.

Another sense in which the model is flexible relates to its independence from any specific construal of the aims of science in general, of explanation, or inter-theoretic reduction, and other central questions in philosophy of science. The model functions as a basic descriptive framework and provides a way to describe instances of activities involved in scientific practice in an intelligible way that, importantly, affords many avenues and angles of analysis. Moreover, because of this, the model does not impose a relevance criterion for what should be considered acceptable instances of practice and can apply to large-scale research agendas (as in the case of place cell research) as well as to activities such as obtaining research grants. What this means, in practice, is that the model is not limited to treating those cases that would qualify as relevant topics for philosophers of science under the traditional interpretation of philosophy of science that was the object of criticisms from proponents of the Practice Turn. Scholars such as Douglas (2010) and Howard (2003) have lamented that philosophy of science has gone from a diverse discipline that dealt with topics ranging from the traditional epistemology of science to the role of values in scientific practice to a hyper-specialized field that rejects any historical or sociological work as “falling outside” the scope of philosophy of science proper. The components of the problem-solving account that grounds the model are neutral with respect to such questions. For instance, the notion of a constraint only requires that there is some form of limitation imposed by a relevant factor on the search for a solution. Constraints can be epistemic, methodological, social, or value-laden, and examining where constraints come from opens the door to valuable sociological work, the relevance of which is readily apparent in that understanding how and why a problem is constrained in one way or another is an important piece of knowledge for understanding how a particular episode of scientific practice unfolded historically or, in cases where the model is applied to ongoing work, how it might unfold.

Another aspect of the model’s flexibility I want to emphasize here is that it is not tied to one level, or layer, of scientific practice (e.g. it can handle high-level cases such as the historical development of a research agenda as well as small-scale cases such as the operationalization of a given variable within a laboratory). More importantly, the model can help in understanding how small-scale cases relate to more general cases, because of how the distinction between “goals” and “problems” is defined contextually with respect to the aims of the users of the model. One researcher’s sub-problem could be another’s overarching goal, and it should be possible, in principle, to explain how one relates to the other. This essentially solves the potential scope problem that practice-oriented approaches face because of their emphasis on the importance of examining the details of scientific practice.

With respect to the components of the model itself, I believe it is appropriate to point out that by and large, the heavy lifting is done by accounts of human cognition that come from outside of philosophy of science. One motivation for adopting the heuristic account of problem-solving is that it constitutes our best knowledge of how humans tend to reason and solve problems, or at the very least, it has been a fruitful approach. It is perhaps one weakness of the model that much of it rides on whether the account of human cognition it rests on is accurate. Given how central the role of this account is, if it should turn out that the description of human cognition the model relies on is erroneous, then there would need to be substantial work done to see whether the model can be adapted to work with a different account. I do not wish to venture further into the hypothetical, but it is a possibility that needs to be flagged.

At this point, the main drawback of the model is probably obvious: its flexibility is a double-edged sword. Because of how little the model presupposes and because of the range of possible constraints, goals, problems and heuristics that might be relevant, there are a lot of potentially relevant items to take into consideration when characterizing any of these elements of a case. My solution to this potential difficulty may be unsatisfactory to some, but once again I will invoke the importance of context to argue that despite the potential 'combinatorial explosion' of relevant factors, the context of use of the model will likely impose certain limitations on the set of relevant factors. In a sense, this is because philosophers of science themselves are bound by the same cognitive limitations as scientists (see the introduction to Wimsatt 2007 for a detailed discussion of this). Moreover, I believe that having this potential difficulty is worth the advantages of allowing the model to be compatible with a variety of approaches and standpoints. That is, the lack of clear specifications on what is an admissible source of constraints, for instance, allows for the development of a plurality of approaches to the analysis of scientific practice, but one that is ultimately unified by its underlying reliance on the basic problem-solving model I propose here. Thus, the main drawback of the model serves an important meta-theoretical purpose in allowing for a kind of pluralism that can lead to more robust analyses of science[56], in line with the sort of robustness defended by, for instance, Wimsatt (2007, 43-44) and defined as the multiple determination of a character or phenomenon by multiple, independent means.

---

[56] I want to flag that I put this idea forward as a potential avenue for future research, which I discuss more in the concluding chapter of the dissertation.

5.4 Situating the model

At this point, the model has been found to meet the six adequacy conditions specified in chapter 1, and I have discussed the model's main advantages. I now turn to the question of whether the model is an improvement on existing accounts of science in terms of problem-solving and heuristics.

My aim in this section is to situate the dynamic-iterative model in the philosophy of science literature. To this end, I want to compare it to one other approach in particular: Bechtel and Richardson's (2010) analysis of decomposition and localization, which is extended in Bechtel's analysis of the case of place cells in the form of a framework that supports the view that explanation in neuroscience is mechanistic in nature (2014). It is the most developed large-scale analysis of science in terms of problem-solving and heuristics, and as such is the most relevant point of comparison for the model. There are other discussions of heuristics in science, such as Hey's (2014) analysis of how meta-heuristics may prove useful in medical contexts, but those are by and large highly specific and as such do not provide the same level of generality as Bechtel and Richardson's model[57].

Bechtel and Richardson's analysis of decomposition and localization as heuristic strategies begins with their aim to put forward an account of science as a problem-solving enterprise, one that does away with the traditional idealized conceptions of rationality in favor of a naturalistic understanding of how humans tackle problems. They emphasize that "theoretical development is in part an expression of human cognitive style, a consequence of the typical strategies with which human beings attack problems, and the cognitive limitations that make these strategies necessary" (2010, p. 6). This is in direct continuation with the concerns of proponents of the Practice Turn: instead of requiring scientists to adhere to overly demanding normative principles, our analyses of science

---

[57] One approach that is similar to mine in terms of scope and, to some degree, aims, is Darden's analysis of strategies for theory change and theory generation (1991). This being said, by Darden's own admission, "social and cultural factors influence actual scientific developments, but, as a philosopher, my interests focus on scientific reasoning and the development of scientific ideas" (Darden 1991, 5). In light of this, and because Darden's approach is only tangentially related to mine by its overall aims, I will omit any detailed discussion of it, since from the outset it excludes a class of factors that my model handles.

should begin with an accurate description of how it proceeds. Thus, Bechtel's starting point is similar to my own. However, Bechtel and Richardson were also motivated by the emergence of what they call a "new mechanistic philosophy of science", which was a reaction to the poor fit of the biological sciences within the frameworks inherited from logical positivism, in particular the deductive-nomological model of explanation, which posits that explanation boils down to subsuming phenomena under natural laws (Bechtel and Richardson 2010, xvii). Understanding the specifics on Bechtel's problem-solving approach requires an understanding of how central the idea of mechanistic explanation was to its development. I discussed mechanistic explanation in chapter 3, but a refresher seems appropriate here.

Mechanistic explanation is, broadly construed, a framework within which "explaining a phenomenon" amounts to identifying the mechanism responsible for it (Bechtel 2008, 10)[58]. Importantly, mechanistic explanation should be understood as an alternative to the deductive-nomological (D-N) model of explanation, which held that explanations were represented propositionally and were constructed through logical inference. The model faced serious issues, as it failed to capture important features of causality such as temporal organization and direction of causation and was harshly criticized from various directions. Philosophers of science worked on alternative accounts of explanation, but by and large they have seemed rather slow on the uptake with regards to mechanistic explanation[59] (Bechtel, 2008, p. 13). For fields where laws, or even robust causal generalizations, seem hard to come by (such as biology and the life sciences generally), the strategy of explaining via the identification of mechanisms has been widely used for quite some time. Bechtel defines a mechanism as "a structure that performs a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena"

[58] The idea is far from novel and dates back to ancient Greek atomic theories, which underwent a revival of sorts with the work of mechanists such as Descartes and Boyle. Mechanistic explanation has since become commonplace in certain fields, such as biology.

[59] Salmon (1984) did develop a causal-mechanical account of explanation as an alternative to law-based accounts, but which focused on causal explanation rather than mechanisms.

(Bechtel, 2008, p. 13)[60]. Bechtel discusses three features of this characterization in more detail.

First, mechanisms are individuated in terms of the phenomena they are responsible for. Citing Bogen and Woodward (1988), Bechtel (2008, p. 14) defines phenomena as repeatable occurrences in the world. Second, mechanisms are made up of parts (structural components) and operations (processes or changes involving parts). Note that the relevant parts of mechanisms are *working parts*, viz. those involved in the operations (Bechtel, 2008, p. 14; see also Craver, 2009). This means that a given mechanistic explanation may omit some parts if these parts are not involved in producing the target phenomenon. Moreover, mechanisms need to be decomposed, either structurally (into parts) or functionally (into operations). Operations are localized by being assigned to specific parts. Finally, the parts and operations need to be organized appropriately. The mechanism produces a given phenomenon because its parts and operations are linked and organized in a certain way. Moreover, the relevance of organization is not constrained by the boundaries of a given mechanism, as mechanisms are environmentally situated in two ways: in immediate environment (other mechanisms) and in the environment in which the entire system (an organism, say) is situated. Factors at both of these levels, as well as at the lower level of component parts, can influence the functioning of a mechanism, or even be required for the mechanism to perform its function (Bechtel 2008, see also 2009).

What should be obvious from this characterization of mechanistic explanation is that, as noted in chapter 3, mechanistic explanation as conceived by Bechtel relies on two heuristic strategies: decomposition and localization. It is by decomposing a system into mechanisms and a mechanism responsible for a phenomenon into its component parts and by localizing the relevant operations to specific parts that a phenomenon is explained. Bechtel's analysis of the case of place cells (2014) makes the role of decomposition and localization in allowing for the development of mechanistic explanations quite clear, but what should be noted is that from the outset the development of Bechtel and

[60] It should be noted that this characterization is one among others. Glennan (1996) focuses on the properties of the parts, while Machamer et. al. (2000) focus on the metaphysical status of entities and activities (see also Craver, 2009).

Richardson’s, and later Bechtel’s, account of heuristics in science was tied to that of the mechanistic view of explanation.

In contrast, I have made a conscious effort to divorce the dynamic-iterative model from any kind of definite theory of explanation. Indeed, one of the model’s hallmark features and main advantages is that it does not presuppose any particular conception of explanation, instead adopting a general-purpose construal of scientific practice as problem-solving. This is not to say that decomposition and localization are not important, even central, heuristics, and I believe that Bechtel and Richardson’s focus on those strategies is justified and has led to important insights about how explanation functions in the biological sciences. Nevertheless, tying the development of an analysis of science as problem-solving to mechanistic explanation resulted in Bechtel’s analysis of the case of place cells missing the important fact that the representational construal of brain function, while it does support the use of decomposition and localization, was itself a heuristic, that differed from decomposition and localization in important respects. However, Bechtel’s focus on developing an account of heuristics to support the mechanistic view of explanation resulted in his analysis of the case of place cells to be impoverished due to it focusing too much on the role the commitment to representations played in supporting the use of decomposition and localization.

Thus, while the dynamic-iterative model and Bechtel’s approach share important features, especially with respect to their origins in the methodological criticisms of traditional philosophy of science, they differ in that the former does not presuppose that explanation is mechanistic in nature. This is indicative of a deeper difference between the dynamic-iterative model and Bechtel’s which is that the dynamic-iterative model is a much more general framework intended to function as a minimal descriptive framework: it provides a structure within which more specialized investigations can take place. Rather than being directed at the clarification of how explanation works in the life sciences, the dynamic-iterative model is a general-purpose description of science. One with applications that go beyond the philosophical context, as I discuss in the next section. In summary, while Bechtel’s model and mine share a common origin as well as many theoretical leanings, the differences in the scope of application of the models and their commitment to theories of explanation makes the dynamic-iterative model much more flexible.

The dynamic-iterative model's philosophical value can be further illustrated by a cursory discussion of how its analysis of the case of place cells allows for an important conceptual clarification in the debate over the status of neural representations. The consensus in philosophy of mind and in cognitive science is that to explain behavior by appealing to cognitive factors requires appealing to internal states that function as *representations*. Representations are states that stand in for something else, generally some external state of the world that needs to be made available to some part of the cognitive system that doesn't have direct access to it. Representations are, thus, *information-bearing states.* Note that while this feature of representations is uncontroversial, it is also generally considered insufficient for individuating some state as representational – it needs to be supplemented.

In philosophical circles, debates about what makes a representation a "full" or "true" representation often revolve around the notion of *intentionality*, viz. the *aboutness* of representations. Because the property of "representing" depends on a state or entity standing in for something else, or being about something else, intentionality can be "derived" quite easily: a stop sign on the road stands in for the imperative to stop your car, but the way in which it can be said to "represent" is derivative at best: it is only because you can interpret it as representing a legal imperative (and because this content was imparted to it by social and legal convention) that it *has* the content "stop the car". On certain views (e.g. Haselanger et al. 2003, Ramsey 2007), the sense in which a stop sign can be said to be representational is explanatorily irrelevant, because if such things are granted the status of "full" representations, then the notion becomes theoretically toothless: everything can be a representation (Haselanger et al. 2003, 17). Another version of this worry is expressed in the dynamicist challenge to representationalism. Van Gelder (1995), in particular, develops an argument based on the example of the Watt Governor, a system that performs a function for which representations are usually thought to be required, but that cannot be said to be a representational system[61]. Importantly, the argument relies heavily on the intuition that something as simple as a Watt Governor could not be said to be a representational system in any meaningful sense because it is a “mere” mechanical device. The argument is thus that representations are useless from an explanatory standpoint because it is possible to explain the functioning of the Watt Governor without ever using the notion of representation. Bechtel (2001) argues, however, that there is one interpretation of the notion of representation on which the Watt Governor *is* a representational system, and, crucially, that it is the very same notion of representation that is operative in behavioral and cognitive neuroscience. Bechtel argues that the simplicity of the mechanism (mis)leads Van Gelder to reject the representational interpretation of the Watt Governor. On such an interpretation, the Watt Governor is representational insofar as things such as the present and desired speeds of the flywheel, as well as the operations of closing and opening the valve are represented by the position of the arms and certain rules apply to these representations such that the system produces a certain behavior (opening or closing the valve to attain a consistent speed). Van Gelder maintains that the Watt Governor’s functioning can be explained without ever invoking representations, and he takes this to indicate that the cognitivist approach is flawed because representations are superfluous explanatory constructs – they are not required to provide an explanation of how behavior is produced. The argument here seems to rely on the intuition that a mechanism as simple as the Watt Governor couldn’t reasonably be called “representational”[62].

---

[61] A Watt Governor is a mechanism designed to accomplish a simple task: regulate the output of steam from a steam engine so that the flywheel rotates at a constant speed regardless of the resistance generated by the appliances connected to it. Watt attached a spindle to the flywheel and attached arms to the spindle which, because of centrifugal force, would open out in proportion to the speed of the flywheel’s rotation. A mechanical link between the arms and the steam valve caused the valve to close when the wheel turned too fast, ensuring a consistent steam output from the valve. The system thus performs a form of behavior regulation, but without the use of representations.

[62] Some responses to Van Gelder’s argument agree with this intuition and challenge the argument on different grounds. For instance, Clark and Toribio (1994) argue that genuinely cognitive tasks are “representation-hungry” and that while the Watt governor isn’t representational, other systems may function on similar principles and rely on representational properties to fulfill certain tasks. In a similar vein, Grush (1997) argues that representations are necessary when the system has no immediate access to information it needs to perform its functions, and thus requires a stand-in state that represents that information.

These types of arguments have been highly influential on the literature concerning neural representation. Ramsey (2007) argues that the only sense in which neural activity can be said to be representational is in the minimal sense discussed by Bechtel, which makes the recourse to representational idioms a mere "gloss" imposed on data without substantive importance. Another angle is taken by Cao (2012) when she argues that a "genuine" representation requires an organism-level receiver and that talking about the activities of individual neurons as representational is an inadequate use of the notion.

The analysis of the case of place cells conducted in chapter 4 provides a different way to engage this debate. Instead of establishing a philosophical definition of "representation" and checking whether the notion, as it is used in neuroscience, conforms to that definition, the model allows for the identification of a separate notion of representation, one that is close to the one Bechtel (2011) proposes. Moreover, it makes clear the reasons *why* that notion of representation became central to research on place cells, and the role the notion played, and continues to play, in supporting the development of accounts of spatial cognition. These reasons provide a way to evaluate whether the notion of representation is doing the job scientists need it to do. It may well be the case that the notion of representation used in neuroscience is wrong from a philosophical point of view, but be that as it may, the fact remains the notion plays a crucial role in neuroscience, and the dynamic-iterative model allows for an appreciation of that role, one that may not be as obvious without the descriptive framework provided by the model.

Beyond this example, the model can also be used to put various practice-oriented studies of science in context and relate them to one another. For instance, the analysis of the role that the notion of representation plays in research on place cells can be related to the analysis of the specific experimental challenges encountered in the course of studying how place cells represent place information. The latter is likely to engage with different aspects of practice, but can nevertheless be related to the more general analysis through the structure of problem-solving activities. I now turn to a discussion of the model's potential applications outside of philosophy of science itself.

5.5 Conclusion

Following the application of the model to a real case, I was able to identify potential objections to the model and responding to those objections allowed me to revise some of the model's core concepts. Importantly, this led to the development of the contextual distinction between goals and problems, which itself is now an important feature of the model, one that allows it to not only meet the six adequacy conditions outlined in chapter 1, but also one that bolsters the model's main strength, which is its flexibility.

In section 5.3 I showed how the model's flexibility presented many advantages and allowed for the model to not only provide a general description of scientific practice that respects the methodological insights of the Practice Turn, and flagged its two main weaknesses, viz. its reliance on an account of human reasoning that may prove erroneous (although it is unlikely that the entirety of the heuristics research program is plain wrong) and the fact that the model's flexibility means there are many possibly relevant elements to consider when applying it. Importantly, that last weakness allows for a kind of pluralism that can lead to more robust analyses of science, which I take to be a valuable enough possibility that it makes up for the potential difficulty in setting up the model when applying it to a case.

Finally, I compared and contrasted the model with Bechtel an Richardson's, which is the dynamic-iterative model's main 'competitor' insofar as both models rely on similar bases. However, the dynamic-iterative model has a wider descriptive scope than Bechtel and Richardson's model and, as I intend to show in the first section of the next chapter, this allows the model to be applied outside of philosophy of science.

Chapter 6

6 Conclusion

The dynamic-iterative model, as outlined in the previous chapters, remains at a provisional stage. Nevertheless, I have shown it to be a promising tool for philosophers of science interested in understanding scientific practice. That said, my aim is also to develop a model of scientific practice that can play a role in furthering the social engagement of philosophy of science. As noted in chapter 1, philosophy of science, despite its huge potential social value, is not in a state that is conducive to the realization of this value. Thus, I begin this final chapter with a discussion of the potential application of the dynamic-iterative model as a pedagogical tool for teaching philosophy of science to non-philosophers. I then discuss avenues for future work, focusing on integrating the dynamic-iterative model within the larger context of values-conscious epistemology.

6.1 A practical application

My aim in this section is to argue that the usefulness of the dynamic-iterative model goes beyond its use in philosophical analysis. I develop a tentative formulation of a teaching strategy based on the dynamic-iterative model.

One of the core motivations for this project is that of overcoming the challenge posed by the relative lack of impact philosophy of science has outside of its disciplinary boundaries, a state of affairs discussed in the introduction. As a quick reminder, the issue is that, as Douglas notes, while "philosophers of science have much to offer scientists and the public, I am skeptical that much can be gained by philosophers importing off-the-shelf discussions from philosophy of science to science and society" (Douglas 2010, 317). On her view, this state of affairs can be traced back to arguments in favor of narrowing the scope of topics that should be thought of as philosophy of science in the 1950's, which likely found traction because of the political climate of the time, in particular pressures to depoliticize academia, which led to the discipline focusing exclusively on the epistemic aspects of science (Douglas 2010, 321; Reisch 2005). The consequence was that the diversity of topics discussed in discipline's flagship journal all but disappeared by the 1960's (Douglas 2010, 321; Howard 2003). This disciplinary narrowing also led to the development of research problems and concerns that were peculiar to philosophers of science and in particular to the entrenchment of the "traditional" approaches criticized by proponents of the Practice Turn (see chapters 1 and 2 for a more detailed discussion of this topic).

This kind of disciplinary isolation is not peculiar to philosophy of science. Indeed, there are strong indicators that philosophy as a whole has become isolated from other academic fields, and that this isolation has serious practical implications ranging from a negative perception of philosophy not only in the academic world at large, but also in the more constrained sphere of the humanities (Higgins and Dyschkant 2014), to the attitude of granting agencies such as the United States' National Science Foundation (NSF) towards interdisciplinary work involving philosophy (Tuana 2013).

That philosophy overlaps with other disciplines is nothing new. Interactions between ontology and physics or philosophy of mind and psychology would appear to be obvious and expected, insofar as these are cases where philosophers and non-philosophers are ostensibly talking about the same things. Higgins and Dyschkant assert that there are two attitudes that have been adopted by philosophers with regards to these overlaps: integrationism holds that the traditional methods of philosophy should be augmented by those of other disciplines that are relevant to their objects of study, while autonomism insists on the autonomy and authority of philosophy (Higgins and Dyschkant 2014, 372). Bealer proposes two theses that capture the authority and autonomy principles: most central questions of philosophy can be answered by philosophical investigation without substantive reliance on the sciences (autonomy) and in cases where the answers of philosophy conflict with those of science (save for a few exceptions)[63], the authority of philosophy is in principle greater (Bealer 1996, 121). While many philosophers would recognize that integrationism is preferable to isolationism, it appears that few practice it. Higgins and Dyschkant observe that an analysis of 142 493 citations from 4727 texts on

---

[63] These exceptions are, specifically, cases that involve the definition of semantically unstable terms like "water" (Higgins and Dyschkant 2014, 376).

ontology and related topics reveals that philosophers by and large cite very little sources outside of philosophy (Higgins and Dyschkant 2014, 377). Moreover, they found that two of the most influential papers dealing with ontological questions (Ashburner et al. 2000, from the perspective of biology and Gruber 1993, from the perspective of computer science), each with over 10 000 citations, have been engaged by “strikingly few” authors (Higgins and Dyschkant 2014, 376)[64]. As a final illustration, they note that one of the most widely cited authors within the philosophical ontology network, Ned Markosian, is accused by Ladyman et al. (2007) of defending an account that would make most of the entities posited by fundamental physics, including the universe itself, non-physical. That is not an issue in and of itself. The issue is the failure, on the part of Markosian, to even engage with the theories his account undermines (Higgins and Dyschkant 2014, 377).

On Higgins and Dyschkant’s view, this isolationism leads to practical consequences tied to the perception of philosophy as stagnant and irrelevant (2007, 373). Coupled with the difficulties faced by philosophers of science in making meaningful contributions outside of philosophy of science even when their input would be invaluable, finding avenues to pull philosophy out of its isolation is an important project. One way to achieve this (although by no means the only way) is to examine the way philosophy is taught and devise better teaching strategies. At least as far as the potential impact of philosophy of science outside of philosophical circles goes, there seems to be value in devising better pedagogical approaches that would focus on the practical uses of philosophy of science in scientific contexts. Philosophers of science often lament the fact that despite the invaluable potential contributions philosophy of science could make, there is very little actual collaboration between philosophers and scientists. For instance, Tuana argues that integrating methods from the humanities, particularly philosophy, into interdisciplinary work in the sciences can reveal value-laden decisions embedded in research models that would otherwise go unnoticed (Tuana 2013, 1956). This being said, integrating philosophy in science education is not an easy task.

---

[64] It should be noted here that there is an argument to be made that the two papers referenced by Higgins and Dyschkant use “ontology” in a different sense than philosophers tend to use it. Nevertheless, the lack of interaction between philosophers and non-philosophers remains a concern.

This is due in part to disciplinary differences with regards to skill acquisition: philosophy tends to focus on deep reading of core debates, while science teaching relies on learning conventional *methods*, that is, how to *do* certain things (Grüne-Yanoff 2014). The dynamic-iterative model can be worked into a scaffolded teaching strategy (Padgett Walsh et al. 2014; Concepción 2014) that emphasizes the skills that philosophers of science bring to the table. This would not make science students philosophers of science, but would hopefully make them philosophically literate science students, and could be a way to foster engagement between science and philosophy. I intend to argue that using a simplified version of the dynamic-iterative model can bring to light unexamined assumptions in various areas of scientific practice, and that, in general, it can ground a method for getting science students to ask the kinds of questions philosophers of science are concerned about.

Pedagogy has not been a core concern of philosophers of science, at least not recently. It is nevertheless possible to discern a general trend in how the topic is treated in the literature. By and large, there is little attention paid to teaching *methods*. Bradner (2015) emphasizes the development of more inclusive syllabi that represent a variety of normally underrepresented perspectives instead of discussing, for instance, a few feminist authors at the end of the semester. This is an important step not only for teaching philosophy of science but for teaching philosophy in general. The teaching method itself remains somewhat unexamined, however, as Bradner recommends the usual approach of close reading of texts relating to core issues in philosophy of science. Contrast this with active learning exercises such as that proposed by Hardcastle and Slater (2014), consisting of having students try to determine the contents of a closed container without opening it, which offers a way to discuss myriad core issues in the philosophy of science while having students engage in an activity highlights the relevance of these issues. Similarly, textbooks designed to help teach philosophy of science to science students, such as Johansson's undergraduate textbook (2015) do not go beyond presenting traditional debates and ideas in a slightly less technical way than if the textbook was geared towards philosophy undergraduates, as reveals a cursory examination of the contents based on Boersema’s review of the textbook (2017).

A more serious issue is that some philosophers construe the process of teaching philosophy of science to scientists as a way to remedy what they deem are liabilities inherent to science education. For instance, Grüne-Yanoff argues that the fact science education relies on teaching “mere” conventional methodology is an epistemic liability because students are not taught to go beyond these conventions (Grüne-Yanoff 2014, 118). While this is true, there are reasons why conventional methodology is central to science education: the activities involved in scientific practice are complex and demanding both in terms of time and cognitive resources. That science students are not taught to question the implicit assumptions that ground the methods they are taught is an issue, but solving it is unlikely to happen by requiring that in addition to the methods, student learn all of the underlying assumptions and theoretical foundations. Nevertheless, Grüne-Yanoff identifies three important obstacles to teaching philosophy of science to science students. First, there are administrative and disciplinary boundaries to overcome in that teaching students from other departments may prove difficult given the obligations professors have to their home departments. More importantly, however, are the obstacles of scepticism towards philosophy and the possible lack of motivation of science students to take philosophy of science courses (Grüne-Yanoff 2014, 116). The teaching strategy I develop below is meant to overcome these last two obstacles.

6.1.1 A novel teaching strategy

To begin with, here is a condensed version of the argument I made above:

P1 – Philosophy of science is conceptually/practically useful for scientists, users of science and the general public.
P2a – Philosophy of science is insular and thus cannot be easily applied outside its disciplinary context.
P2b - Teaching approaches in philosophy of science either **(a)** embrace insularity or **(b)** treat science condescendingly, leading to lower chances of reaching scientists.
C1 - Current “state” of philosophy of science and teaching strategies will not achieve P1.

My aim here is to propose a teaching strategy that 1) shirks insularity, 2) treats scientific practice with respect without "deferring to the wisdom" of scientists, 3) is tuned to the aim of P1 and 4) informed by current best practices in pedagogy.

The core idea of this teaching strategy is to get science students to engage in the actual work of philosophy of science as quickly as possible instead of guiding them through the philosophical canon. That is, the strategy is to *use* philosophy to analyze science as early as possible in the semester. This is where the dynamic-iterative model comes in. As discussed in previous sections, one of the main advantages of the model is that it is relatively economical, conceptually speaking. That is, the model relies on two concepts that are likely to be familiar to most science students: problems and heuristics. Both of these concepts can be defined easily and illustrated with myriad examples both from science and from everyday life. The same is true of important peripheral concepts such as constraints, tractability, and so on. Thus, the first step is to introduce students to the basic conceptual framework: when scientists conduct research, they encounter problems, and to help them solve those problems they apply cognitive strategies known as heuristics. The literature on problem-solving and heuristics offers myriad resources for introducing students to these notions, one example being Kahneman's *Thinking, Fast and Slow* (2011), which contains accessible formulations of these key notions.

In the context of using the model as a pedagogical tool, I believe it is preferable to employ a simplified version which focuses on the notion of a *research question*, that is, a notion flexible enough to capture a range of aspects of scientific practice from the general goal of a research agenda to the specific goal pursued in an experimental paper. Students are then shown examples of how problems arise when trying to answer a research question, and how these problems are solved by the researchers, what strategies they used to solve them, and how the solution to these problems influenced the data acquired and how this data led to revisions to the research question posed initially. This imparts to the students the idea of a dynamic process while showing the importance of being aware of underlying assumptions and potential biases in the strategies employed. The simplified model is represented in figure 7 below.

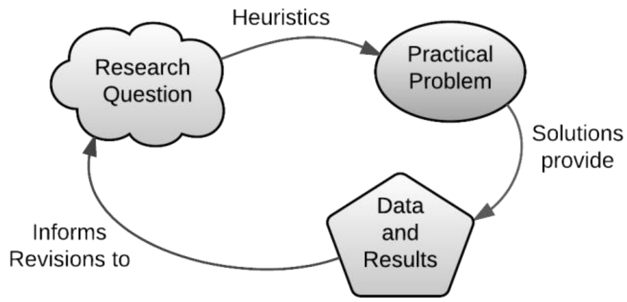


Figure 7: Simplified dynamic-iterative model

After being introduced to the conceptual framework, students *apply* the framework to a case of *their* choice. This ensures that students are more likely to be motivated, in addition to potentially reducing their workload should they be able to apply the framework to a case they are currently engaged in (for instance, many honors students work in laboratories, and could thus pick as a case for the course something they are currently working on in the laboratory). As an example, take the case of recording neural activity to determine how the brain processes 3D objects (based on Snow et al. 2011):

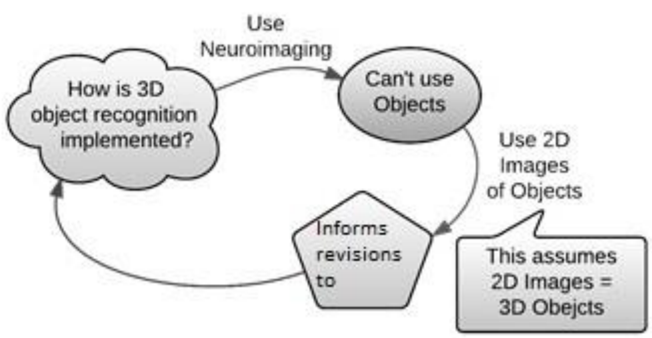


Figure 8: Simplified model filled in with details of example case

There is a research question, and then a choice made to use neuroimaging to answer the question. There is a practical problem, which is that scanners do not afford the use of real objects because of restricted space, and thus the choice is made to use images of objects as a stand-in for real objects. Students are led to tease out the underlying assumption made here and then asked to consider what might happen if the assumption is erroneous, or whether there are good reasons to assume 2D images are processed in the same way as 3D objects. In doing so, they begin to engage in philosophical work without needing an extensive knowledge of the philosophical canon.

The second component of the teaching strategy is the use of a scaffolded teaching method, which involves 1) discrete steps (requiring breaking down assignments in specific tasks), 2) a progressive sequence (the steps are organized in a sequence such that each subsequent step builds on the previous in an explicit manner) and 3) specific interventions between steps (timely and actionable feedback). The main advantages of scaffolded instruction are threefold. First, it helps avoid “cognitive overload”: instead of having students write an entire term paper, the assignments build towards the completion of a case study over the course of the semester, allowing students to focus on one task at a time (Padgett Walsh et al. 2014, 484). Second, it allows students to receive “formative assessments” during the writing process. That is, students are provided with feedback at various stages instead of being given a term paper with comments and feedback that cannot be applied since the grade has already been registered. According to Padgett Walsh et al., this allows students to become familiar with the process of making revisions to a paper that go beyond small-scale edits, a crucial skill in most, if not all, academic domains (Padgett Walsh et al. 2014, 485). Finally, scaffolded instruction provides students with a higher chance for success, allowing them to complete a complex project in discrete and attainable stages (Padgett Walsh et al. 2014, 485). Other advantages of the scaffolded instruction model are that it is highly adaptable and, crucially, emphasizes practice-based learning in that it relies on a cycle of assignments and feedback.

There are four main advantages to using this strategy to teach philosophy of science to science students. First, their understanding of philosophical work is framed in terms of science they are familiar with. Second, the problem-solving framework and the scaffolded instruction structure make it possible to maximize the time spent *doing* philosophical work: the notion of problem-solving is both simple and familiar to many STEM students, and the structured assignments provide a foundation for a final project they are likely interested in completing because the project itself is relevant to their interests *as science students*. Third, the strategy allows students to practice philosophy of science and at the same time demonstrates the value philosophy of science can have in the context of practicing science. To go back to the 3D object recognition example, once students are led to recognize the assumption underlying the production of data, they can be introduced to the relevant philosophical literature on the topic, in this case readings in the philosophy of experiment (e.g. Hacking 1983), questions about validity and reliability (e.g. Sullivan 2009) or data-phenomena inferences (Bogen and Woodward 1988). Finally, the "case based" approach allows for flexibility in course design: the strategy can be used to teach to STEM students from a variety of disciplines.

The two components of this teaching strategy are independent from one another. Indeed, it might be the case that the problem-solving model that provides the basic conceptual framework is erroneous, or otherwise unsuited to pedagogical contexts, which would not affect the relevance of scaffolded instruction. This being said, the dynamic-iterative model is uniquely suited to a theory-light teaching strategy that focuses on imparting the skills associated with a philosophical background, in large part because of its small set of key concepts and overall simplicity. The model itself can become fairly complex in some cases, but it does not need to be. Its adaptability to different contexts is an advantage in terms of allowing for diverse and relevant practical applications, such as the one outlined here. This is something that other models would struggle to accomplish, and while the dynamic-iterative model is not necessarily developed as a teaching strategy, it is developed as a flexible framework that is likely to be useful in practical contexts. Another advantage of the model is its independence from any specific theory of explanation, or of the aims of science. The model really is nothing more and nothing less than a basic descriptive framework. I believe that this shows that the model has the potential to be adapted to be useful in many different contexts.

6.2 Future work

The dynamic-iterative model is, at its core, a straightforward descriptive framework. Its descriptive scope differs from that of Bechtel and Richardson's problem-solving and heuristics model in that it can account for a wider range of activities and strategies, as demonstrated in chapters 4 and 5. This is in part because the latter is developed for the purpose of grounding the mechanistic view of explanation in the life sciences, while the former is developed for the purpose of providing an accurate and flexible descriptive framework on top of which analyses of scientific practice can be built. I want to insist on the importance of understanding the *motivations* that underlie the development of a descriptive framework. A key motivation for the development of the dynamic-iterative model was that of making progress towards realizing philosophy of science's potential impact factor on science and society. That is, my aim was to ensure that this new descriptive model could be put to use in the context of direct engagement and collaboration between philosophers of science and users and practitioners of science. The dynamic-iterative model is thus best understood as a basic descriptive framework intended as a flexible tool for engaging scientific practice. The model supports the development of analyses of science (as exemplified in chapter 4) as well as the development of practical, applied work (exemplified in the previous section). In the remainder of this concluding chapter, I want to discuss other possible applications as well as avenues for future work.

The potential philosophical applications of the model go beyond the sort of conceptual clarification work hinted at in chapter 5. Indeed, one of the most important features of the model is its flexibility with respect to the specification of relevant constraints on problem-solving. In the case study conducted in chapter 4, I focused on epistemic constraints (e.g. the constraint imposed by behaviorist frameworks that explanations of spatial learning and navigation in rats should not invoke anything like internal cognitive states) and other constraints traditionally considered "internal" to scientific practice[65]. While the lack of a clear discussion of the role non-epistemic values may play in the development of research agendas in these pages is unfortunate, I believe that the dynamic-iterative model might offer a useful framework for thinking about how such values influence scientific practice. Specifically, because the model does not impose relevance conditions on constraints, both epistemic and non-epistemic values are candidates for sources of constraints. The model is thus compatible with the idea that the

---

[65] What I mean here is that while such epistemic constraints are part of the traditional vocabulary of philosophy of science as constraints internal to science, I want to flag that the constraints I discuss below are also part of science itself, although they were not considered to be on traditional accounts.

distinction between epistemic and non-epistemic values is not a hard and fast one and that both sorts of values have important implications for the development of scientific knowledge (Longino 1996; Rooney 1992). Moreover, the connections between the study of heuristics and the study of implicit bias (e.g. Beaulac and Kenyon 2014) offers a way to think about ideological bias in science, with the notion of constraint acting as a conceptual tool that allows one to trace the specific impact a given bias may have on the development of a research agenda. While demonstrating that the dynamic-iterative model can integrate such concerns would require more applied work, I believe that, at least in principle, the model is compatible with such work.

This brings me to another important aspect of the dynamic-iterative model: its potential us for pluralistic analyses of scientific practice. For instance, an analysis of the case of place cells that focused on different kinds of constraints than those discussed in chapter 4 could reveal different facets of the development of place cell research. Again, this is a possibility afforded by the flexibility of the model, and “pluralistic” analyses can happen in various ways. One way is to analyse the same instance of practice but to focus on different constraints or, rather, sources of constraints. Another way is to recontextualize a research problem into a research goal and over iterative analyses, construct a more complete understanding of a research agenda. For instance, the analysis in chapter 4 did not focus on the experimental process – it merely identified some experimental problems, but a different analysis of the same episode might take one of these experimental problems and analyze how the solutions to that problem shaped the development of the research agenda. Science is a dynamic process, and the possibility of applying the model at different levels while making the relationship between those levels explicit, and a part of the analysis, allows for a more complete appreciation of this dynamic nature.

The important thing here is that the model is not limited to the analysis of conceptual problems, and can take the practical problems that come up during experimentation as a starting point, or as an overarching goal, and produce an analysis of the reasons that guided the choice of such and such experimental technique, or protocol. Similarly, the choice of data analysis techniques can be the focus of analysis, and the model provides a way to relate these choices to higher-level theoretical problems. This flexibility does mean that the model does not come “fully assembled”, but, again, the model is intended as a basic descriptive framework within which analyses of scientific practice can be conducted. The specific aims and purposes of the analysis will dictate how the model is applied, at which level of practice it is applied, what the relevant sources of constraints are going to be, and so on.

One might wonder what, if any, is the advantage of such a model. I defended the model’s usefulness in chapter 5, in particular with respect to its descriptive scope: the model is developed independently of any other substantive project such as an account of explanation and as such is not limited to those heuristics relevant for the construction of, say, mechanistic explanations, whereas Bechtel and Richardson’s 2010 model (which is also applied in Bechtel’s 2014 analysis of place cell research and is developed for the purpose of grounding the mechanistic view of explanation in the life sciences) is. This difference in scope comes down, in part, to a crucial difference in the motivations behind the two models. Again, both my model and Bechtel and Richardson’s are motivated by a rejection of “traditional” philosophy of science, especially with respect to conceptions of rationality. The difference is that my model is also motivated by the project of making philosophy of science more directly relevant to users and practitioners of science.

I outlined one possible application in section 6.1, viz. the use of the model as the basis for a scaffolded teaching strategy. This idea highlights an important aspect of the model, which I have mentioned before: the fact it is nothing over and above a basic descriptive framework. In the case of the pedagogical application of the model, the model itself only provides the conceptual background for the teaching strategy, that is, it does not, by itself, provide a fully-fledged approach to teaching philosophy of science. This also goes for the philosophical applications of the model: the analysis of the case of place cells shows that the model can provide a workable description of a case, but that capturing some aspects of the dynamics of the scientific process might require supplementing the model with other approaches. As an example, I have mentioned that it would be possible, in theory, to recontextualize some of the practical problems involved in conducting the experimental work involved in solving the empirical problems that made up the place cell research agenda. While there are no conceptual obstacles to this, it should be noted that the model, as I have presented it here, does not have all the necessary tools to do this: experimental aspects of scientific practice are contained in the category of practical problems and I have not provided a detailed analysis of how such problems relate to conceptual problems or empirical problems. Nevertheless, it is possible to see how the model might be supplemented by other approaches. For instance, Sullivan's analysis of experimental protocols in neuroscience (2009) provides a framework through which the practical problems involved in the place ell research agenda could be better understood. This being said, bringing in complementary accounts is not always a straightforward matter.

Future work on the dynamic-iterative model will partially thus focus on evaluating those areas where the model needs to be supplemented by other approaches and how these other approaches can be integrated with the dynamic-iterative model. Again, the model itself is a very powerful descriptive tool, but this descriptive power is largely a result of the model's flexibility and of the relative conceptual economy that characterizes it. This comes at the price of, perhaps, needing more "assembly" than traditional analytical tools in philosophy of science. Nevertheless, I believe the gains in descriptive accuracy and the accompanying potential for meaningful engagement with practitioners and users of science are valuable enough to warrant the adoption of the dynamic-iterative model.

Bibliography

Akins, Kathleen. "Of Sensory Systems and the 'Aboutness' of Mental States." *The Journal of Philosophy* 93, no. 7 (1996): 337–72. https://doi.org/10.2307/2941125.

Andersen, Hanne, and Brian Hepburn. "Scientific Method." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2016., 2016. http://plato.stanford.edu/archives/sum2016/entries/scientific-method/.

Ankeny, Rachel A. "Fashioning Descriptive Models in Biology: Of Worms and Wiring Diagrams." *Philosophy of Science* 67, no. 3 (2000): 272.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, et al. "Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium." *Nature Genetics* 25, no. 1 (May 2000): 25–29. https://doi.org/10.1038/75556.

Barkow, Jerome, Leda Cosmides, and John Tooby. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press, 1992.

Bealer, George. "A Priori Knowledge and the Scope of Philosophy." *Philosophical Studies* 81, no. 2–3 (March 1, 1996): 121–42. https://doi.org/10.1007/BF00372777.

Beaulac, Guillaume, and Tim Kenyon. "Critical Thinking Education and Debiasing (AILACT Essay Prize Winner 2013)." *Informal Logic* 34, no. 4 (December 10, 2014): 341–63. https://doi.org/10.22329/il.v34i4.4203.

Bechtel, William, and Robert C. Richardson. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Reissue. The MIT Press, 2010.

Belnap, Nuel D., and T. B. Steel. *The Logic of Questions and Answers*. First Edition edition. New Haven: Yale University Press, 1976.

Bickle, John. *Philosophy and Neuroscience*. Dordrecht: Springer Netherlands, 2003. http://link.springer.com/10.1007/978-94-010-0237-0.

———. "Reducing Mind to Molecular Pathways: Explicating the Reductionism Implicit in Current Cellular and Molecular Neuroscience." *Synthese* 151, no. 3 (2006): 411–434.

Bishop, Michael A., and J. D. Trout. *Epistemology and the Psychology of Human Judgment*. Oup Usa, 2005.

Boersema, David. "Philosophy of Science for Scientists, by Lars-Göran Johansson; and The Nature of Scientific Knowledge: An Explanatory Approach, by Kevin McCain." *Teaching Philosophy* 40, no. 3 (2017): 385–389.

Bradner, Alexandra. "How to Teach Philosophy of Science." *Teaching Philosophy* 38, no. 2 (2015): 169–192.

Cao, Rosa. "A Teleosemantic Approach to Information in the Brain." *Biology & Philosophy* 27, no. 1 (January 1, 2012): 49–71. https://doi.org/10.1007/s10539-011-9292-0.

Chang, Hasok. *Inventing Temperature: Measurement and Scientific Progress*. Oup Usa, 2004.

Chirimuuta, M. "Extending, Changing, and Explaining the Brain." *Biology and Philosophy* 28, no. 4 (2013): 613–638.

Chow, Sheldon J. "Many Meanings of 'Heuristic.'" *The British Journal for the Philosophy of Science* 66, no. 4 (2015): 977–1016. https://doi.org/10.1093/bjps/axu028.

Clark, Andy, and Josefa Toribio. "Doing Without Representing." *Synthese* 101, no. 3 (1994): 401–31.

Clarke, Murray. *Reconstructing Reason and Representation*. 1st ed. MIT Press, 2004.

Concepción, David. "Engaging Novices: Transparent Alignment, Flow, and Controlled Failure." In *Philosophy Through Teaching*, edited by K. Hermberg, R. Kraft, and E. Esch, 129–36. American Association of Philosophy Teachers, 2014.

"Connectome - About the CCF (CCF Overview)." Accessed November 12, 2017. https://www.humanconnectome.org/about-ccf.

Costandi, Mo. “Anti-Connectome-Ism | Mo Costandi.” *The Guardian*, September 21, 2012, sec. Science. http://www.theguardian.com/science/neurophilosophy/2012/sep/21/connectome-review.

Craddock, R. Cameron, Saad Jbabdi, Chao-Gan Yan, Joshua T. Vogelstein, F. Xavier Castellanos, Adriana Di Martino, Clare Kelly, Keith Heberlein, Stan Colcombe, and Michael P. Milham. “Imaging Human Connectomes at the Macroscale.” *Nature Methods* 10, no. 6 (June 2013): 524–39. https://doi.org/10.1038/nmeth.2482.

Craver, Carl F. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press, Clarendon Press ;, 2007.

———. “Prosthetic Models.” *Philosophy of Science* 77, no. 5 (2010): 840–851.

Craver, Carl F., and Lindley Darden. “Discovering Mechanisms in Neurobiology: The Case of Spatial Memory.” In *Theory and Method in Neuroscience*, edited by P. K. Machamer, Rick Grush, and Peter McLaughlin, 112–137. Pittsburgh: University of Pitt Press, 2001.

Cross, Ryan. “Neuroscientists Can Now Map Any Human Brain That Comes Their Way.” MIT Technology Review, 2016. https://www.technologyreview.com/s/601940/the-map-of-the-human-brain-is-finally-getting-more-useful/.

Darden, Lindley. *Theory Change in Science: Strategies From Mendelian Genetics*. Oxford University Press, 1991.

Davis, Tyler, and Russell A. Poldrack. “Measuring Neural Representations with FMRI: Practices and Pitfalls.” *Annals of the New York Academy of Sciences* 1296 (August 2013): 108–34. https://doi.org/10.1111/nyas.12156.

Douglas, Heather. “Engagement for Progress: Applied Philosophy of Science in Context.” *Synthese* 177, no. 3 (September 30, 2010): 317–35. https://doi.org/10.1007/s11229-010-9787-2.

Feest, Uljana. “What Exactly Is Stabilized When Phenomena Are Stabilized?” *Synthese* 182, no. 1 (September 1, 2011): 57–71. https://doi.org/10.1007/s11229-009-9616-7.

Feigl, Herbert, and May Brodbeck, eds. *Readings in the Philosophy of Science*. First Edition. New York: Appleton-Century-Crofts, 1953.

Fraassen, Bas C. Van. *The Scientific Image*. Oxford University Press, 1980.

Gelder, Tim van. "What Might Cognition Be If Not Computation?" *Journal of Philosophy* 92, no. 7 (1995): 345–81.

Gigerenzer, Gerd. "On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky.," 1996. http://psycnet.apa.org/psycinfo/1996-01780-008.

Giunti, Marco. "Hattiangadi's Theory of Scientific Problems and the Structure of Standard Epistemologies." *British Journal for the Philosophy of Science* 39, no. 4 (1988): 421–439.

Godfrey-Smith, Peter. "Signal, Decision, Action." *Journal of Philosophy* 88, no. 12 (1991): 709–22.

———. *Theory and Reality: An Introduction to the Philosophy of Science*. University of Chicago Press, 2003.

Gould, S. J., and R. C. Lewontin. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme." In *Conceptual Issues in Evolutionary Biology*, 73–90. The Mit Press. Bradford Books, 1994.

Griffiths, Paul E. *What Emotions Really Are: The Problem of Psychological Categories*. University of Chicago Press, 1997.

Gruber, Thomas R. "A Translation Approach to Portable Ontology Specifications." *Knowledge Acquisition* 5, no. 2 (June 1, 1993): 199–220. https://doi.org/10.1006/knac.1993.1008.

Grüne-Yanoff, Till. "Teaching Philosophy of Science to Scientists: Why, What and How." *European Journal for Philosophy of Science* 4, no. 1 (January 1, 2014): 115–34. https://doi.org/10.1007/s13194-013-0078-x.

Grush, Rick. “The Architecture of Representation.” *Philosophical Psychology* 10, no. 1 (1997): 5–23.

Hacking, Ian. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. 1 edition. Cambridge Cambridgeshire ; New York: Cambridge University Press, 1983.

———, ed. *Scientific Revolutions*. Oxford University Press, 1981.

Haraway, Donna. “Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective.” *Feminist Studies* 14, no. 3 (1988): 575–599.

Hardcastle, Gary, and Matthew H. Slater. “A Novel Exercise for Teaching the Philosophy of Science.” *Philosophy of Science* 81, no. 5 (December 1, 2014): 1184–96. https://doi.org/10.1086/678240.

Haselager, Pim, A. de Groot, and H. van Rappard. “Representationalism Vs. Anti-Representationalism: A Debate for the Sake of Appearance.” *Philosophical Psychology* 16, no. 1 (2003): 5–23.

Hattiangadi, J. N. “The Structure of Problems, Part I.” *Philosophy of the Social Sciences* 8, no. 4 (1978): 345–365.

———. “The Structure of Problems, Part II.” *Philosophy of the Social Sciences* 9, no. 1 (1979): 49–76.

Higgins, Andrew, and Alexis Dyschkant. “Interdisciplinary Collaboration in Philosophy.” *Metaphilosophy* 45, no. 3 (2014): 372–398.

Howard, Don A. “Two Left Turns Make a Right: On the Curious Political Career of North American Philosophy of Science at Midcentury.” In *Logical Empiricism in North America, Minnesota Studies in the Philosophy of Science,* edited by A. Richardson and G. Hardcastle, XVIII:25–93. Minneapolis: University of Minnesota Press, 2003.

Jabr, Ferris. “The Connectome Debate: Is Mapping the Mind of a Worm Worth It?” Scientific American, 2012. https://www.scientificamerican.com/article/c-elegans-connectome/.

Jensen, Robert. “Behaviorism, Latent Learning, and Cognitive Maps: Needed Revisions in Introductory Psychology Textbooks.” *The Behavior Analyst* 29, no. 2 (2006): 187–209.

Johansson, Lars-Göran. *Philosophy of Science for Scientists*. 1st ed. 2016 edition. New York, NY: Springer, 2015.

Kahneman, Daniel. *Thinking, Fast and Slow*. Toronto: Doubleday Canada, 2011.

Kaplan, David Michael, and Carl F. Craver. “The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective.” *Philosophy of Science* 78, no. 4 (2011): 601–627.

Kelly, Clare, Bharat B. Biswal, R. Cameron Craddock, F. Xavier Castellanos, and Michael P. Milham. “Characterizing Variation in the Functional Connectome: Promise and Pitfalls.” *Trends in Cognitive Sciences* 16, no. 3 (March 1, 2012): 181–88. https://doi.org/10.1016/j.tics.2012.02.001.

Kuhn, Thomas S. *The Structure of Scientific Revolutions (2nd Ed.)*. Vol. xiv. Chicago, IL, US: University of Chicago Press, 1970.

Latour, Bruno, and Steve Woolgar. *Laboratory Life: The Construction of Scientific Facts*. Edited by Jonas Salk. 2nd ed. edition. Princeton, N.J: Princeton University Press, 1986.

Laudan, Larry. “A Problem-Solving Approach to Scientific Progress.” In *Scientific Revolutions*, edited by Ian Hacking. Oxford University Press, 1981.

———. *Progress and Its Problems: Towards a Theory of Scientific Growth*. First edition. Berkeley, Calif.: University of California Press, 1978.

Lloyd, Elisabeth A. “Evolutionary Psychology: The Burdens or Proof.” *Biology and Philosophy* 14, no. 2 (1999): 211–33.

Lloyd, Elisabeth A., and Marcus W. Feldman. “Evolutionary Psychology: A View from Evolutionary Biology.” *Psychological Inquiry* 13, no. 2 (January 1, 2002): 150–56. https://doi.org/10.2307/1449175.

Lombrozo, Tania, Anastasia Thanukos, and Michael Weisberg. “The Importance of Understanding the Nature of Science for Accepting Evolution.” *Evolution: Education and Outreach* 1, no. 3 (June 20, 2008): 290–98. https://doi.org/10.1007/s12052-008-0061-8.

Longino, Helen E. “Can There Be A Feminist Science?” *Hypatia* 2, no. 3 (1987): 51–64.

———. “Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy.” In *Feminism, Science, and the Philosophy of Science*, edited by Lynn Hankinson Nelson and Jack Nelson, 39–58. Kluwer Academic Publishers, 1996.

Magnus, P. D. “Underdetermination and the Problem of Identical Rivals.” *Philosophy of Science* 70, no. 5 (2003): 1256–1264.

Mandik, Pete. “Varieties of Representation in Evolved and Embodied Neural Networks.” *Biology and Philosophy* 18, no. 1 (January 1, 2003): 95–130. https://doi.org/10.1023/A:1023336924671.

Marcus, Gary. *Kluge: The Haphazard Evolution of the Human Mind*. Reprint. Mariner Books, 2009.

Mayo, Deborah G. “Novel Evidence and Severe Tests.” *Philosophy of Science* 58, no. 4 (1991): 523–552.

McCain, Kevin. *The Nature of Scientific Knowledge: An Explanatory Approach*. Springer, 2016.

McMullin, Ernan. “Laudan’s Progress and Its Problems.” *Philosophy of Science* 46, no. 4 (December 1, 1979): 623–44.

Millikan, Ruth G. *Language, Thought and Other Biological Categories*. MIT Press, 1984.

Morgan, Joshua L., and Jeff W. Lichtman. “Why Not Connectomics?” *Nature Methods* 10, no. 6 (June 2013): 494–500. https://doi.org/10.1038/nmeth.2480.

Nagel, Ernest. *The Structure of Science: Problems in the Logic of Scientific Explanation*. Harcourt, Brace & World, 1961.

Newell, Allen, J.C. Shaw, and Herbert A. Simon. “The Processes of Creative Thinking.” In *Contemporary Approaches to Creative Thinking: A Symposium Held at the University of Colorado*, edited by H. E. Gruber, G. Terrell, and M. Wertheimer, 63–119. The Atherton Press Behavioral Science Series. New York, NY, US: Atherton Press, 1962.

Nickles, Thomas. “Scientific Problems and Constraints.” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1978 (1978): 134–148.

———. “Theory Generalization Problem Reduction and the Unity of Science.” In *PSA 1974*, edited by R. S. Cohen, C. A. Hooker, A. C. Michalos, and J. W. Van Evra, 33–75. Boston Studies in the Philosophy of Science 32. Springer Netherlands, 1976. http://link.springer.com/chapter/10.1007/978-94-010-1449-6_3.

———, ed. *Thomas Kuhn*. Contemporary Philosophy in Focus. Cambridge, U.K. ; New York: Cambridge University Press, 2002.

———. “What Is a Problem That We May Solve It?” *Synthese* 47, no. 1 (1981): 85–118.

Nordmann, Alfred. Review of *Review of* Science after the Practice Turn in the Philosophy, History, and Social Studies of Science, by Léna Soler (eds.) Sjoerd Zwart, Michael Lynch, and Vincent Israel-Jost, June 21, 2015. http://ndpr.nd.edu/news/58957-science-after-the-practice-turn-in-the-philosophy-history-and-social-studies-of-science/.

O’Keefe, J., and D. H. Conway. “Hippocampal Place Units in the Freely Moving Rat: Why They Fire Where They Fire.” *Experimental Brain Research* 31, no. 4 (April 14, 1978): 573–90.

O’Keefe, J, and J Dostrovsky. “The Hippocampus as a Spatial Map. Preliminary Evidence from Unit Activity in the Freely-Moving Rat.” *Brain Research* 34, no. 1 (November 1971): 171–75.

O'Keefe, J., and M. L. Recce. "Phase Relationship between Hippocampal Place Units and the EEG Theta Rhythm." *Hippocampus* 3, no. 3 (July 1993): 317–30.

O'Keefe, John. "Spatial Cells in the Hippocampal Formation." presented at the Nobel Prize Award Ceremony, Stockholm, Sweden, December 7, 2014.

O'Keefe, John, and Lynn Nadel. *The Hippocampus as a Cognitive Map*. Oxford; New York: Clarendon Press ; Oxford University Press, 1978.

Padgett Walsh, Kate, Anastasia Prokos, Sharon R. Bird, and Philosophy Documentation Center. "Building a Better Term Paper: Integrating Scaffolded Writing and Peer Review." Edited by Michael Goldman. *Teaching Philosophy* 37, no. 4 (2014): 481–97. https://doi.org/10.5840/teachphil201410225.

Penfield, Wilder, Theodore Rasmussen, and National Institute on Drug Abuse. *The Cerebral Cortex of Man: A Clinical Study of Localization of Function*. New York: Macmillan, 1950.

Polya, George. *How to Solve It*. Garden City, New York: Doubleday, 1957.

Popper, Karl R. *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge, 1962.

———. *The Logic of Scientific Discovery*. Routledge, 1959.

Ramsey, William M. *Representation Reconsidered*. Cambridge University Press, 2007.

Reisch, George A. *How the Cold War Transformed Philosophy of Science: To the Icy Slopes of Logic*. Cambridge University Press, 2005.

Richardson, Robert C. *Evolutionary Psychology as Maladapted Psychology*. A Bradford Book, 2010.

Rooney, Phyllis. "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?" *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992 (1992): 13–22.

Rouse, Joseph. "Kuhn's Philosophy of Scientifc Practice." In *Thomas Kuhn*, edited by Thomas Nickles, 101–22. Cambridge, U.K. ; New York: Cambridge University Press, 2002.

Schenk, Thomas. "No Dissociation between Perception and Action in Patient DF When Haptic Feedback Is Withdrawn." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 32, no. 6 (February 8, 2012): 2013–17. https://doi.org/10.1523/JNEUROSCI.3413-11.2012.

"Science Aims to Explain and Understand." Accessed October 12, 2016. http://undsci.berkeley.edu/article/0_0_0/whatisscience_04.

Shae, Brendan. "Popper, Karl: Philosophy of Science | Internet Encyclopedia of Philosophy." Accessed October 12, 2016. http://www.iep.utm.edu/pop-sci/.

Shapin, Steven, and Simon Schaffer. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton, N.J: Princeton University Press, 1985.

Smith, Stephen M., Peter T. Fox, Karla L. Miller, David C. Glahn, P. Mickle Fox, Clare E. Mackay, Nicola Filippini, et al. "Correspondence of the Brain's Functional Architecture during Activation and Rest." *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 31 (August 4, 2009): 13040–45. https://doi.org/10.1073/pnas.0905267106.

Soler, Léna, Sjoerd Zwart, Michael Lynch, and Vincent Israel-Jost. "Introduction." In *Science after the Practice Turn in the Philosophy, History, and Social Studies of Science*, edited by Léna Soler, Sjoerd Zwart, Michael Lynch, and Vincent Israel-Jost, 123–51. New York: Routledge, 2014.

———, eds. *Science after the Practice Turn in the Philosophy, History, and Social Studies of Science*. New York: Routledge, 2014.

Sporns, Olaf, Giulio Tononi, and Rolf Kötter. "The Human Connectome: A Structural Description of the Human Brain." *PLOS Computational Biology* 1, no. 4 (September 30, 2005): e42. https://doi.org/10.1371/journal.pcbi.0010042.

"SPSP." Accessed April 20, 2017. http://www.philosophy-science-practice.org/en/mission-statement/.

Sullivan, Jacqueline A. "Reconsidering 'Spatial Memory' and the Morris Water Maze." *Synthese* 177, no. 2 (November 1, 2010): 261–83. https://doi.org/10.1007/s11229-010-9849-5.

———. "The Multiplicity of Experimental Protocols: A Challenge to Reductionist and Non-Reductionist Models of the Unity of Neuroscience." *Synthese* 167, no. 3 (2009): 511–539.

Sullivan, Jacqueline Anne. "Construct Stabilization and the Unity of the Mind-Brain Sciences." *Philosophy of Science* 83, no. 5 (December 2016): 662–73. https://doi.org/10.1086/687853.

Talbot, S. A., and W. H. Marshall. "Physiological Studies on Neural Mechanisms of Visual Localization and Discrimination *." *American Journal of Ophthalmology* 24, no. 11 (November 1, 1941): 1255–64. https://doi.org/10.1016/S0002-9394(41)91363-6.

Thornton, Stephen. "Karl Popper." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016., 2016. http://plato.stanford.edu/archives/win2016/entries/popper/.

Tolman, Edward C. "Cognitive Maps in Rats and Men." *Psychological Review* 55, no. 4 (1948): 189–208. https://doi.org/10.1037/h0061626.

Tooby, John, Leda Cosmides, and H. Clark Barrett. "Resolving the Debate on Innate Ideas: Learnability Constraints and the Evolved Interpenetration of Motivational and Conceptual Functions." In *The Innate Mind: Structure and Contents*. New York: Oxford University Press New York, 2005.

Tuana, Nancy. "Embedding Philosophers in the Practices of Science: Bringing Humanities to the Sciences." *Synthese* 190, no. 11 (July 1, 2013): 1955–73. https://doi.org/10.1007/s11229-012-0171-2.

Van Essen, D. C., K. Ugurbil, E. Auerbach, D. Barch, T. E. J. Behrens, R. Bucholz, A. Chang, et al. "The Human Connectome Project: A Data Acquisition Perspective." *NeuroImage*, Connectivity, 62, no. 4 (October 1, 2012): 2222–31. https://doi.org/10.1016/j.neuroimage.2012.02.018.

Van Essen, David C., Stephen M. Smith, Deanna M. Barch, Timothy E. J. Behrens, Essa Yacoub, and Kamil Ugurbil. "The WU-Minn Human Connectome Project: An Overview." *NeuroImage*, Mapping the Connectome, 80, no. Supplement C (October 15, 2013): 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041.

Viger, Christopher D. "Is the Aim of Perception to Provide Accurate Representations? A Case for the 'No' Side." In *Contemporary Debates in Cognitive Science*. Malden MA: Blackwell Publishing, 2006.

Wimsatt, William C. "Aggregate, Composed, and Evolved Systems: Reductionistic Heuristics as Means to More Holistic Theories." *Biology and Philosophy* 21, no. 5 (2006): 667–702.

———. "Reductionism and Its Heuristics: Making Methodological Reductionism Honest." *Synthese* 151, no. 3 (2006): 445–475.

———. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press, 2007.

Wood, Diana F. "Problem Based Learning." *BMJ* 326, no. 7384 (February 8, 2003): 328–30. https://doi.org/10.1136/bmj.326.7384.328.

Woodward, James. "Scientific Explanation." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2014., 2014. http://plato.stanford.edu/archives/win2014/entries/scientific-explanation/.

Woody, Andrea I. "Chemistry's Periodic Law: Rethinking Representation and Explanation After the Turn to Practice." In *Science after the Practice Turn in the Philosophy, History, and Social Studies of Science*, edited by Léna Soler, Sjoerd Zwart, Michael Lynch, and Vincent Israel-Jost, 123–51. New York: Routledge, 2014.

Wright, Jessey. “The Analysis of Data and the Evidential Scope of Neuroimaging Results.” *British Journal for the Philosophy of Science*, forthcoming.

Curriculum Vitae

Name:	Frédéric-Ismaël Banville
Post-secondary Education and Degrees:	Université de Montréal Montréal, Québec, Canada 2007-2010 B.A.
	Université du Québec à Montréal Montréal, Québec, Canada 2010-2012 M.A.
	The University of Western Ontario London, Ontario, Canada 2012-2018 Ph.D.
Honours and Awards:	Bourses d'excellence de l'UQAM pour les cycles supérieurs (FARE) 2011-2012
	Social Science and Humanities Research Council (SSHRC) Joseph-Armand Bombardier Canada Graduate Scholarship 2012-2015
	Ontario Graduate Scholarship 2015-2016
Related Work Experience	Research Assistant Université du Québec à Montréal 2010-2011
	Teaching Assistant The University of Western Ontario 2013-2016
	Research Assistant The University of Western Ontario 2016

Publications:
Banville, F.-I. (2012) "*Implementing Revisionism: Assessing a Revisionist Theory of Moral Responsibility*", Ithaque, 10, p. 115-135