Title
Methods in Historically-Informed Philosophy of Science

Author
Mitsch, Chris

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Methods in Historically-Informed Philosophy of Science

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Logic & Philosophy of Science

by

Chris Mitsch

Dissertation Committee:
Professor James Weatherall, Chair
Chancellor's Professor Jeffrey A. Barrett
Distinguished Professor Emeritus Penelope Maddy

2020

DEDICATION

To

my family, broadly construed

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my committee: Jim Weatherall, for advocating for me; Pen Maddy, for teaching me how (not) to do philosophy; and Jeff Barrett, for early support of my investigation of von Neumann's axiomatization of quantum mechanics. Additionally, thanks to my other advancement committee members: Jeremy Heis, for his help as I learned German and showing me what good historical work looks like, and Ginny Richards, for paying me to press buttons when I needed time away from LPS. Thanks also to Cailin O'Connor and J.B. Manchak, for making me feel welcome.

I would like to thank the administrative staff in LPS and the School of Social Sciences, especially Patty Jones and John Sommerhauser, for all of their hard, uncelebrated work.

I would like to thank my friends and colleagues for all their help and support along the way. In particular, I would like to thank Alex Hodge-Wallis, Stella Moon, Nathan Richards, Sarita Rosenstock, Jeffrey Schatz, Will Stafford, Alan Webb, and Kino Zhao for their moral support.

Finally, I would like to thank: my parents, for making my education possible and believing in me; my brother, for making exploration enticing; and my partner, Nicole, for everything. I can't (and won't) imagine who I would be without you, Nicole.

Chapter 3 of this dissertation was made possible in part through the support of grant #61048 from the John Templeton Foundation, as well as grant support from the University of California, Irvine, School of Social Sciences. The opinions expressed therein are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

Chapter 1 is published in *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*: 'An Examination of Some Aspects of Howard Stein's Work' by Chris Mitsch, 2019, reproduced by permission of Elsevier https://doi.org/10.1016/j.shpsb.2019.04.001.

VITA

Chris Mitsch

Ph.D in Logic & Philosophy of Science **2020**
University of California, Irvine *Irvine, California*

M.A. in Logic and Philosophy of Science **2017**
University of California, Irvine *Irvine, California*

B.A. in Logic **2013**
Ohio State University *Columbus, Ohio*

A.A. in Fine Arts **2012**
Edison Community College *Piqua, Ohio*

ABSTRACT OF THE DISSERTATION

Methods in Historically-Informed Philosophy of Science

By

Chris Mitsch

Doctor of Philosophy in Logic & Philosophy of Science

University of California, Irvine, 2020

Professor James Weatherall, Chair

While the history of science has received increased attention from philosophers of science in recent decades, the methods with which we import the former into the latter are less well-understood. This dissertation investigates two common import strategies and discussions of one significant historical question to this end. The first chapter addresses Howard Stein's "Yes, but...," which is typically conceived of as an entry into the scientific realism debate. To the contrary, I show that it denies the methodological errors that allow the debate to take place. The second chapter addresses attempts to derive norms for scientific practice from science's history. Examining a paradigmatic example, I show that extant arguments of this sort fail to establish any norms because they fail to establish any causal relationship. The third chapter addresses the history of von Neumann's so-called no hidden variables theorem. By situating it in the context of Hilbert's axiomatic method I show that, contrary to most historical accounts, von Neumann's theorem was sound.

Introduction

Here I investigate ways of integrating science's history into its modern, practicable philosophy. In this vein, I consider two approaches commonly used in philosophy, delineate methodological problems often encountered therein, and articulate alternative methods. In a final chapter, I discuss an historical incident that demonstrates how distorting anachronisms can be of our history.

In Chapter 1, I consider the use of science's history as the arbitrator for disputes among extant positions in philosophical debates. I show that this is the role science's history was expected to play in the scientific realism debate of Putnam, Laudan, and Boyd, whose central question was whether our best scientific theories are "true" in addition to being effective. Both the scientific realist and instrumentalist claim science's history confirms their view, and I show that there is a sense in which each is correct.

However, echoing Stein, I argue that this stalemate is evidence that the history has been misused. Indeed, I show that the debate itself only makes sense when one makes two serious historiographical errors, namely, assuming that science can and should be discussed in such general, clear, and certain terms as the philosopher's "theory" and "true," and that these terms apply universally throughout time and place. Absent these assumptions, I argue that the debate should rather concern how (psychologically, physiologically) humans come to the scientific understandings they do.

In Chapter 2, I consider the use of science's history as a generator of norms for contemporary scientific practice. A paradigmatic example is Stanford's argument for scientific instrumentalism on the grounds that there exist scientifically-serious unconceived alternatives to our best scientific theories, which he claims is evident from science's history. According to Stanford, science's history shows that instrumentalists are better able to discover unconceived alternatives. For this reason, he claims that scientists should be instrumentalists.

I argue that Stanford has given no reason to believe that instrumentalists are better able to discover unconceived alternatives. I show that his claim that instrumentalists outperform realists is a fundamentally causal claim and that, for this reason, he must provide evidence of causation. Nevertheless, by contrasting Stanford's argument with causal arguments that use established methods, I show that his argument provides no evidence of causation. I conclude by suggesting that he failed to generate a norm for contemporary scientific practice not because of anything about the history, but instead because the realism–instrumentalism debate is predicated on a poorly-motivated abstraction from actual scientific practice. To the contrary, I suggest that this abstraction be abandoned, at least at present.

In the final chapter, I reconsider the history of von Neumann's axiomatization of quantum mechanics, particularly of his so-called no hidden variables proof. Accounts to date have considered this proof in the context of later work by Bohm and Bell. In this later work, a hidden variables interpretation is characterized precisely by denying certain assumptions of von Neumann's. Consequently, the reception of von Neumann's proof has been overwhelmingly negative insofar as von Neumann wrongly concludes from it that a hidden variable theory of quantum phenomena is impossible.

To the contrary, I argue that the proof was the natural, and highly-informative, conclusion of von Neumann's use of the axiomatic method. Drawing on a characterization of Hilbert's axiomatic method that emphasizes the importance of a unique mathematical representation for physical theories, I show that von Neumann's axiomatization achieved its goal of "ordering" quantum mechanics and "orienting" future research. Indeed, I suggest that it had done this already in 1927, years before the proof was published as part of von Neumann's 1932 textbook. Crucial to my argument is the recognition that 'quantum mechanics' referred specifically to the statistical interpretation of the transformation theory. Consequently, when von Neumann claimed that hidden variables could not be added to quantum mechanics, he was not claiming that *no* hidden variable theory of quantum phenomena was possible. Thus, the negative reception of von Neumann's work relies on a false representation of his work.

Chapter 1

An Examination of Some Aspects of Howard Stein's Work

1.1 Introduction

Some who read Stein's "Yes, but..." consider his remark that there is "no difference that makes a difference" between realism and instrumentalism a reflection of the paper's most important lesson for those engaged in the debate. Stanford, for instance, focuses on Stein's suggestion that "the dispute between realism and instrumentalism is not well joined," in that there is a convergence of ambitions between a sophisticated realism and a sophisticated instrumentalism (Stanford, 2005, 404-5). With this lesson in mind, the "no difference" comment seems a natural slogan for the paper's central takeaway.

I don't think the meaning of this slogan is properly understood, however. I will argue that the keys to understanding this remark are not found in the relationships among the various doctrines of realism and instrumentalism *per se*; rather, the keys are found to be methodological morals that actually *preclude* the debate. These methodological morals concern the relationship between, on the one hand, philosophy (of science) and science and, on the other, philosophy of science and the history of science. I will show that Stein's view on the contemporary realism-instrumentalism debate (RID) is better seen as a corollary to his views on these other matters than as an isolable verdict.[1] As such, a proper understanding of his views on the RID must appreciate this broader context. To set this context, I will draw heavily from Stein's dissertation. While it is less clear that they are to form a coherent document, one will find also in his later work considerable support for what I argue here. However, the fact that these are evidently consistent with his dissertation provides support for the idea that his work forms a fairly coherent whole.

The structure of this paper is as follows. In §2.1 I will argue that Stein is wary of the distortions and dangers of undue generalizations, certainty, and clarity in the history and philosophy of science, and that this is alluded to in the opening paragraph of "Yes, but…". Then I will sketch the two methodological morals as they appear in his earlier work: that scientific inquiry is a dialectic (§2.2) and that it is an enterprise (§2.3). The conclusions that (I suggest) follow most directly from the former concern what may be called the relation of philosophy (of science) to science, whereas those following most directly from the latter concern the relation of the history of science and the philosophy of science. (Though this structure suggests that the two morals, and perhaps the areas in which the conclusions following from them tend to fall, are more-or-less distinct, I have imposed this structure primarily for readability.) Finally, I will sketch a reading of "Yes, but…" (§3). In §3.1, I highlight the centrality of the two morals, especially the dialectical conception, to the argument. Then, I show how RID does not respect these morals and is thereby either irrelevant to (§3.2) or wrong of (§3.3) scientific inquiry. Lastly in §3.4, and beyond its irrelevance or inaccuracy, I show how these morals suggest that RID is a distraction from a

[1]It is at the very least clear that his view of the RID did not change significantly after 1982 when he made reference to the talk that later became the paper (Stein, 1982, 569-70) (Stein, 1987, 391-2).

legitimate inquiry that is of relevance to scientific inquiry. I will conclude (§4) by providing an explanation of Stein's remark that there is "no difference that makes a difference".

1.2 Methodological Morals

1.2.1 Explicit reflection on principles: generality, certainty, clarity

My first observation is, I hope, not contentious: relative to much work in general philosophy of science, the ideas Stein tends to engage with are often not (yet) clear. Perhaps this is just a matter of style, his arguments and conclusions still possessing the usual implicational force. I don't think this is correct, however. My suggestion is this: this is a feature of his work because (§2.1) he is consciously wary of the distortions and dangers of undue generalizations, certainty and clarity, (§2.2) he strongly disapproves of the rhetorical ends to which these have tended to be put in inquiry itself, and (§2.3) likewise disapproves of the ways these have typically been generated from the history. It is, I am suggesting, a reflection of a conscious attention to methodology that is evident throughout his work. In this subsection I will argue that the first part of this suggestion is plausible and is relevant to understanding "Yes, but…".

The title and first paragraph of "Yes, but…" allude especially to his wariness of generalizations:

> By the word "skeptical" [in the title, "…Some Skeptical Remarks on Realism and Anti-Realism"] I do not mean to suggest, primarily, disbelief; my ideal skeptic is Socrates, not Pyrrho. Among the claims put forward in recent years in the name of "scientific realism" there are many things I agree with; but there is also an admixture of what seems to me unclear in conception, or unconvincingly argued.

> This is not so much different from what I thought, in my student days, about the doctrines of logical empiricism—which have since been pretty harshly dealt with. In the latter proceeding, I believe that some rather valuable philosophical lessons have been (at least partly) lost or obscured; and I fear that unless a sufficient ferment of Socratic skepticism is cultured within the realist brew, it will go stale and its vogue too will soon pass.

I note three things here. First is the nature of his skepticism, which is not to suggest disbelief, he says, but to channel Socrates rather than Pyrrho. This is a common refrain of Stein's. It is important to note here because it tells us the kind of argument we should expect in the following pages. Crucially, the argument is dialectical rather than didactic: it gives his realist/instrumentalist interlocutor space to defend their view, and in so doing Stein aims to expose and scrutinize their goals and presuppositions. This implies that we should take him seriously when he says he agrees with many of the individual *claims* put forward in the name of "scientific realism" while having serious concerns regarding its *conception* and *arguments in its favor*. This means, too, that we should neither expect nor impose a clean categorization of 'scientific realism' or 'instrumentalism'; we as observers are invited to suspend any concern for the RID in favor of learning some "rather valuable philosophical lessons [that] have been (at least partly) lost or obscured." In a sense, we should be prepared for the ground under the feet of the RID to shift.

Second, he points to some of his past thoughts that may serve as a rubric for understanding the argument here. The thoughts here are, he says, similar to those he had of logical empiricism in his student days. Though it is obviously a primary concern of his dissertation, there is curiously little explicit reflection on logical empiricism there. For example, the best we may conclude from the introduction is that, insofar as logical empiricism was to set philosophy on a "safe path" to becoming a science or put it on a "scientific" footing, its failure was evident (Stein, 1958, 2;10). However, what *is* present is a (large) collection of *particular* observations regarding philosophy, science, and their histories. His introduction to the dissertation is dedicated to telling us why his focus is so narrow.

Insofar as his dissertation has a clearly-enunciated aim, it is to determine the "role of philosophy itself in relation to its subject matters," and in particular "its relationship to the sciences" (Stein, 1958, 1-2). To do this, one could "survey and attempt to systematize" the extant doctrines and then provide a "philosophy of philosophy" that makes cogent the "diversity of philosophic positions on the subject of philosophic diversity." However, the results of such a program are tainted by several doubtful presuppositions. Of major concern is that "[s]uch a program obviously presupposes the greatest possible, that is to say the amplest and also the deepest and most accurate, erudition." Such an approach also presumes that philosophy *has* an intelligible role (that can be determined by philosophical reflection) and that there *is* a pattern in the history that will provide insight into the nature of philosophy (Stein, 1958, 3). There are also two more practical concerns. First, it is clear that "success is contingent upon the adequacy of the sample" (Stein, 1958, 3, fn. 1), which (I suspect) is here even more difficult to guarantee than in more mundane situations. Second, the history of philosophy provides some grounds for skepticism. Regardless of their truth, that so many philosophers "have analyzed and counterbalanced the doctrine of other philosophers in order to educe true principles," is quite persuasive on this point. "That is," he continues, (Stein, 1958, 4)

> the very prevalence in history of essays of the sort contemplated must inevitably tend to the result that another such attempt, *even if it succeed* [sic], will be likely to experience the fate of those of the past. We have undergone the sophistication of diversity; and any claim, on the part of a philosophy, to be either scientific or encyclopedic, can expect to be entertained with a knowing smile.

This pushes him to find some more modest, promising route of attack. Stein suggests an alternative route to insight "by way of a reflective examination of particular problems" instead of by attempting to provide a systematic philosophy of philosophy (Stein, 1958, 7-8). This, of course, comes with its own difficulties and dangers:

> If one intends to examine problems of the sciences, for instance, one had best be clear as to the stringency of the demands which those particular Muses place upon their votaries; many a distinguished toe has been stubbed for too lightly tripping upon their exigent terrain. So much for difficulty; as for defect, the proceeding is open to the very grave charge, on the philosophical side, that what passes for a plain historical method is rather likely to let principles go by default, or more precisely to move on a ground insufficiently criticized and understood; that is to say, (1) failure to establish first principles first may lead judgment to be based unwittingly upon prejudices; (2) principles inadequately understood may be incorrectly applied, producing merely specious conclusions—conclusions whose contraries may equally be "demonstrated"; (3) an inquiry whose principles are insufficiently established and codified may, even though correctly conducted at each particular stage, lead to changes in the foundation and so to overall inconsistency as it progresses.

For this reason, we must explicitly reflect on what guides our inquiry. Here the sciences suggest some optimism, for they have faced similar difficulties and dangers yet, historically, have at least in part overcome them (Stein, 1958, 10). We are thus encouraged to study how this is so in the sciences; this Stein does with respect to mechanism from Galileo and Huygens to Schrödinger and Bohr. And as he does consider there to be problems of principle comparable to those of philosophy, some light is shed on the problem in philosophy. But, to reiterate what was stated above, there are few "general" claims made in his dissertation, and the conclusions that are present are sometimes less precise than evocative. Luckily, I am not the only one who thinks so. The closing paragraph of his dissertation alludes to this regarding specifically his final thoughts on Schrödinger's view of human mentality (Stein, 1958, 397):

> The immediately foregoing remarks are extremely disorganized, and it is very hard to attach to some of them any precise meaning. In the latter respect, the problem that they center upon is quite comparable to the problem of the ultimate nature of force as it was discussed in the seventeenth century. That discussion was far from fruitless (although the problem was never solved, but ultimately dissolved into far more complicated ones). The point to be made about them is that such remarks ought neither to be dressed up until they can pass for a system of philosophy, nor argued out of existence because of the difficulty of clarifying them: it is philosophical, if not indeed wise, to recognize that in our age as in ages past there are some things that we know; some things that we are more or less clear about not knowing; and some things about which even clarity cannot yet be achieved.

Of course, this does not mean that nothing is clear in Stein or that he believes this is so; he is no obscurantist, nor, I think, does he take himself for one. It does, however, show that he is comfortable with a lack of clarity when no better is possible at the time, and that this has been successful in the history of science; unclarity is an inevitable consequence of seeking wisdom, one that should not be avoided.

Third, and picking back up on the analogy between his thoughts on realism and logical empiricism, the lessons of logical empiricism concern not just the doctrine itself. The crux of the matter is embodied in Stein's later recounting of the discussion following a colloquium presentation of Quine's. Stein was encouraged that Quine and Carnap agreed on the conditions by which their disagreement could be settled, but distressed about the *way* Quine continued to engage (making it out to be a matter of intelligibility rather than fruitfulness). In fact, Stein seems to side with Quine rather than Carnap on several of the major doctrinal points separating the two. For instance, he tells us: that "we really do not have a satisfactorily analyzed epistemological 'basis' for any department of knowledge, mathematics and logic included" (Stein, 1992, 283); that the analytic-synthetic distinction "serves little purpose"; and that the observational-theoretical distinction, as Carnap envisioned it, does not work (likewise dooming his inductive logic) (Stein, 1992, 291). Moreover, it seems clear that moving in the dialectical direction Stein suggests clearly entails a much more idiosyncratic understanding of theory-world relations, one unlikely to be captured adequately by anything so precise as the frameworks Carnap envisioned (Stein, 1992, 292). This view of Carnapian doctrine is present in Stein's earlier work, too, e.g. (Stein, 1970, 286-7). There, too, the concern is less about doctrine and more about method; his assessment of the "danger" pointed to is one (he says) Carnap would also agree with, despite Carnap's framework seeming to give rise to it. Like in Stein (1992), Stein is singing the praises of reserve and discrimination:

> No attempt to delimit, systematically and globally, the procedures and notions that are empirically legitimate. . . has really succeeded.To say this. . . is to deprecate the appeal to programmatic notions as if the program had been realized: this leads to specious criticism.. . . It has been possible for scientists, in creating, criticizing, modifying, and revolutionizing their theories, to apply what is valid in [principles such as "hypothesis non fingo" and the verifiability theory of meaning], despite the lack of an adequate precise general formulation. There is no reason why philosophers of science cannot do the same.

Thus it should be clear that a scholastic salvaging of doctrinal details is not Stein's aim. Rather, what Stein hopes to have persuaded us of—"that there is more in [Carnap's] philosophy than most current representations of it imply" (Stein, 1992, 294)—appears primarily *methodological* in character. In the end it was Carnap who was the good fallibilist, not

Quine: Carnap, he says, was open to wholesale change of his approach to understanding scientific activity (including the doctrine he actually expounded), should it prove not fruitful or representative; Quine, despite his professedly leaving "all open to the flow of experience," obstinately insisted that his current formulation was definitive (Stein, 1992, 292).[2] Beyond the obvious *thematic* similarities with his dissertation (to be noted), this understanding fits well with the few remarks there directed at logical empiricism (above, and one to follow).

On this basis, I hope it will be granted that the first part of my suggestion is plausible: Stein is consciously wary of the distortions and dangers of undue generalizations, certainty and clarity.

1.2.2 Genuine skepticism and the dialectic

In this subsection I will argue that the second part of the above hypothesis is plausible: that Stein strongly disapproves of the rhetorical ends to which undue generalizations, certainty and clarity have historically been put in scientific inquiry. In arguing for this, I will sketch the relationship that he believes has, historically, *properly* obtained between philosophy and scientific inquiry. This will come down to *genuine* skepticism, or, what is a right consequence of it, the dialectical nature of science. The clearest examples of his disapproval concern Descartes and the energeticist/phenomenological school of the late nineteenth century (especially Mach). These two cases feature rather prominently throughout his work, including his dissertation. In the latter, each is set up in contrast to another: Descartes to Galileo, and the energeticist/phenomenological school to the atomists.

[2]This fits well, too, with the following (Stein, 2004, 164):

> a critical mistake has repeatedly been made by philosophers—certainly including natural philosophers. This mistake is the assumption that a clarification of "ideas", or concepts, should always—or *can* always—*precede* the advance of knowledge.... More recently, the logical empiricist school—in its early, and still best-known, positions—had the same view; Carnap's later writings show a quite different view of the matter, and this development on his part is what makes me regard him as... a far deeper philosopher than is generally believed.

Take Descartes and Galileo first. As Stein points out, there is a substantial amount of similarity between the two, for "Descartes, like Galileo, knew the laws of inertia; Descartes, like Galileo, knew that the weight of a body produces an acceleration in free fall"; continuing the comparison, "and yet it is Galileo, and not Descartes, who is the father of the new mechanics" (Stein, 1958, 58). A more direct comparison can be made regarding the case of falling bodies, where again they were similarly poised to make an advancement (Stein, 1958, 63-4). Each began their investigations by (mistakenly) taking the uniformity in the mode of acquisition of increments of velocity to be "in the sense that equal gains of speed are made over equal distances of fall" (Stein, 1958, 64). Koyré, an earlier interpreter of the relation between them, had therefore concluded (in Stein's words) "that the difficulties into which they fall were due to a common, focal sin against time" (Stein, 1958, 82). As it turns out, according to Stein, there *was* a common error across Descartes and *earlier* Galileo; however, contrary to Koyré, the error Galileo makes in his *reductio ad absurdum* of the hypothesis that equal gains of velocity are made over equal spaces traversed has a different basis despite being analogous (Stein, 1958, 75-6).

What ultimately interests Stein is the distinction in method. Each began with the same mistake, yet Galileo, and not Descartes, recognized his error and reached the correct conclusion (Stein, 1958, 82-3). In this way, the case is similar to "the respective roles of Descartes and Galileo in the theory of impact" (Stein, 1958, 83). One might consider pinning the differential success on Galileo's acceptance, and Descartes rejection, of continuity and of relativity—the "two great principles that dominated fundamental theory in physics from the seventeenth century to Einstein" (Stein, 1958, 85). But this isn't quite right: "while it is not *per se* wrong or defective to make what may happen to prove the wrong theoretical commitment, it is defective, it is "unphilosophical" to make any such commitment without adequate reservations or grounds" (Stein, 1958, 85). It is Descartes's "hubristic" claim of "more than moral certainty" for his principles of philosophy that Stein is reacting to.

Descartes' vice, Stein suggests later, was "a tendency to jump too positively to conclusions, a kind of anxiety about attaining final clarity and certainty" (Stein, 1958, 289). In contrast, Galileo (Stein, 1958, 87):

> possesses the related merits of reserve and discrimination—the twin virtues of genuine skepticism; one might say, the Platonic virtues of *sophrosunē* and *dikaiosunē*, of balance and judgment.... Above all, he knew the essential difference between the two attributes of a theory, *looking good*, and *being true*; and he possessed a fine sense of theoretical relevance. Where Descartes pretended to think out clearly the fundamental principles by which to achieve a rational account of the world, and then by applying those principles to achieve that account, Galileo sought to develop adequate principles in the course of the endeavor to deal with the facts of motion...

Though not always named thusly, these twin virtues of genuine skepticism are ever-present themes in Stein's work (compare also, e.g., (Stein, 1974, 397)).

However, Descartes's is not the only way this anxiety for ultimate clarity and certainty has damaged inquiry. For this reason, the damage is properly said to have been caused by *reactions* to the anxiety. Where Descartes reacted with hubris, others have reacted with despair: the anxiety "not only can lead to the neglect, at great cost, of what though important and fruitful is imperfect, but also can produce as reaction that Pyrrhonism or "skeptical despair" which Kant refers to... as the "*euthanasis* of pure reason"" (Stein, 1958, 289). This despair comes in both positive and negative forms. At the outset of the dissertation's body, Stein remarks on an especially strong, and especially negative form. In contrast to Schrödinger—who believes that the "obvious inability of present-day physics and chemistry to account for [the phenomena of life] is no reason at all for doubting that they can be accounted for by those sciences" (Stein, 1958, 11)—"the philosopher may think that he has better reasons than the mere imperfection of present-day physics and chemistry, for doubting that those sciences can account for the phenomena of life" (Stein, 1958, 12-3). "But," Stein says, "this is a dubious road; philosophical refutations of the possibility of science accounting for some domain of phenomena have notoriously been the preludes of scientific conquest" (Stein, 1958, 13). This evaluation of skeptical despair is common throughout Stein's work, e.g., in the more recent (Stein, 2004, 166):

> *I* think—to make one last appeal to the history of philosophy—that the fate of Locke's view that scientific physics is impossible; of Kant's view that scientific chemistry is impossible; of Comte's view that knowledge of the chemistry of the *stars* is impossible; should all conduce to skepticism about that kind of philosophical skepticism.

Skeptical despair in facing the anxiety is damaging when cast in a positive form, too. This is exemplified in the phenomenalism/positivism at the turn of the twentieth century. What makes this example remarkable

> is that the protagonists were not chiefly philosophers, and were not "disillusioned" about the prospects for the advance of science, but were physicists and chemists who espoused with enthusiasm a definite positive program. A second [fact making the example remarkable] is that whereas essentially no contributions to the sciences were in fact made under the aegis of that program, which soon collapsed altogether, the more general philosophical views of its exponents have continued to exert a considerable influence upon the views of both philosophers and scientists to the present day.

The particular case is of the atomic hypothesis. By about 1860, the doctrine of atomic weight was fully established. However, it was still possible "to argue that *no convincing evidence at* *all existed for the hypothesis that matter is composed of discrete ultimate parts*" (Stein, 1958, 293). Indeed, alternate to the atomistic program, there was real reason for the energeticists to hope for a general theory of the modes of energy (Stein, 1958, 295).

However, the energeticists' argument against atoms came not via the evidence but via the perversion of an "economical" manner of speaking. Essentially, they applied different standards to their own and the atomic theory (Stein, 1958, 299). What had previously been a fruitful positivistic tendency in chemical inquiry was, in the hands of those like Mach, combined "with philosophical influences derived from Kant's critique, to produce a school of physical chemists who completely rejected the theory of atoms" (Stein, 1958, 294). This led to, for one, a specious characterization of atoms as "unobservable" for their inability to be observed in the same way as "everyday" objects (Stein, 1958, 299-300). What happened was a "subtle transformation" of the metaphysical doctrine of strict positivism, (Stein, 1958, 302)

> by which *certain* theoretical conceptions have their credentials challenged, while it is forgotten that the doctrines in themselves apply quite as well to other conceptions which pass unchallenged; the reality of atoms, for instance, is denied, but the awareness is suppressed that the particular sense of "reality" employed is such that the reality of chairs is equally—or at least comparably—impeachable.

What therefore follows from even a *positive* skeptical despair is "the tendency to put a stop at a certain stage to the prosecution of theoretical questions by the remark that the aim of science is, after all, just to present the actual facts: description rather than explanation, as the often-quoted dictum of Kirchoff puts it" (Stein, 1958, 303).[3]

[3]Mach in particular equivocates on the word 'sensation' to "dispose wholesale of the problems of physics in relation to biology" (Stein, 1958, 302). He "finally says that it is inappropriate and absurd to seek an explanation of *the natural, i.e., physiological process* of sensation in terms of the conceptions of mechanics" (Stein, 1958, 303). But this temptation—"to consider the physiological facts of sensation as simpler and more immediate than the facts of mechanical motion"—might be forestalled by thinking of the physiology of non-human animals (Stein, 1958, 303).

But there is a simple objection to such a perversion of the (otherwise benign) economical manner of speaking (Stein, 1958, 303-4):

> it is an interesting and defensible view that the object of all science is, in some sense, the fullest possible "description" or "representation" of natural phenomena; but then the fact that the theoretical constructions so abstract and so little obvious as the scheme of Newtonian mechanics and its successors have been of such extraordinary power, and capable of such extension and such great precision in the representation of phenomena, deserves more consideration than it tends to be given in the context.

This economical *manner of speaking* is perfectly acceptable—that is, without an accompanying "subtle transformation." Stein, for instance, is comfortable rephrasing Newton as saying "here is a conceptual scheme involving a certain conception of 'causes' which appears to be eminently useful—even perhaps of ultimate significance—in the elucidation of phenomena" (Stein, 1958, 390). The manner of speaking simply serves to play up a theory's being a conceptual scheme open to revision or replacement or, equivalently, to minimize readers' "reading in" certainty that is not implied. But this is not the same as what I will call Mach's economical *view* (Stein, 1958, 390-1):

> What trouble there is appears to have come from a neglect, by Mach, to consider that the enormous success of Newton's scheme in providing "economical" (and true) representations is itself a fact, not only of history, but about the world: the world is so constituted that the conception, for example, of "internally determined" or natural motions versus "forced" or violent motions leads to a far less economical representation of far fewer natural processes than the conception of "natural powers" as general laws determining the product of mass and acceleration.

So with the economical manner of speaking theories still are about the constitution of the world, whereas with Mach's economical view this has dropped out. Such an economical view of scientific explanation also "led Poincaré to dismiss as idle the question of the *true* mechanical explanation of electromagnetism" (Stein, 1958, 392).

The lesson here is to maintain a healthy balance of the dual aspects of genuine skepticism—to avoid the trappings of an anxiety for clarity and certainty. This he made clear in his concluding remarks (Stein, 1958, 389):

> There are certain lessons that the history of science seems to teach, and an attempt has been made to adumbrate some of them. But perhaps the most important lesson—and here our examination of history has bordered on moralizing!—is the virtue of a certain combination of enthusiasm for the development and consequential application of systems of concepts, with a kind of caution or ultimate skepticism of the ultimacy of ultimate principles and programs.

What is more, this applies also to conceptions of science itself (Stein, 1958, 389):

> In particular, we have seen twice—in Descartes, and in the energeticist and phenomenologist school of the late nineteenth century—the sort of mischief that can be worked by too rigid an adherence to a creed of ultimate clarity and certainty. And the question arises whether the logical empiricist movement in contemporary philosophy, influenced as it has been by such predecessors as Mach, and committed to a sort of clarity and precision in the *language* of science, may not be in danger of working the same sort of mischief.

And here we have a more substantial commentary—and further reinforcement of the above interpretation of Stein's remarks—on logical empiricism. In tying it directly to the failures of Descartes and Mach, he makes it clear that the worry for logical empiricism's future is methodological: if its creed is adhered to too rigidly, the same mischief that was worked by Descartes and Mach could be worked by logical empiricism.

Finally, how an individual maintains a healthy balance brings us to the double-facedness, or dialectical nature, of inquiry.[4] Newton and his contemporaries all appreciated that experience had the final say regarding theories. However, experience should also play a subtler, less well-appreciated role (Stein, 1958, 97-8):

> In the absence of a general guarantee, the actual execution of the program requires, in Newton's conception, the utmost fidelity to detail in studying natural phenomena. This conception is one of Newton's monumental contributions to science: that in the investigation of natural phenomena details are important, that they may provide the clue to important knowledge and that they must be regarded as a touchstone of truth.

Thus, the role of experience for Newton was "not only that of the supreme arbiter, but also that of the only "official" guide" in inquiry (Stein, 1958, 119). Many of his contemporaries, and (I add) many since, have failed to appreciate that nature plays this second role;[5] Huygens and Hooke, for instance, confused inductions from phenomena for hypotheses in arguing against Newton's hypothesis about the constitution of light (Stein, 1958, 98-9) (Stein, a, 15).

In sum, science is dialectical in this thoroughgoing back-and-forth with the world, and it is a proper balancing of the two aspects of genuine skepticism that enables it to function healthily. This requires "the utmost fidelity to details."[6]

[4]See also, e.g., (Stein, 1990b, 38-9).

[5]According to Stein, this seems to be the significance of Newton's rule IV (Stein, 1958, 98-9).

[6]For more on what this fidelity to detail looked like in astronomy post-Newton, I suggest Smith (2014).

1.2.3 The "unphilosophical fallacy" and the enterprise

In this subsection I will argue that Stein also disapproves of the way that undue generalizations, certainty and clarity have been generated from the history of science. Whereas §2.1 addressed historiography more generally, this subsection will address manifestations of a specific, but common, error among historically-minded philosophers: the "unphilosophical fallacy." It will thereby focus on the second moral: the relation of history of science to philosophy of science. I will begin by probing a passage wherein Stein has paused to reflect on his own historical method. I will then briefly draw out of his dissertation a cluster of claims that appear to be consequences of his historical method.

What Stein notes first is a necessity to distinguish between "questions of historical interpretation—even where such questions involve a component that can appropriately be called "philosophical"—from questions of the philosophy of science *per se*" (Stein, 1958, 172). A first-pass understanding might be this: we must distinguish between the philosophy of/in our historical figures and the philosophy of science-as-it-is, which is to say as-we-know-it-to-function. Cognizance of this distinction should, for one, mean we attend especially to actions and reflections of our forebears ways as distinct from a *proper* and *objective* understanding of them.

This first-pass understanding is on the right track, for it is, of course, right to distinguish between our own views and interpretive principles and the view of an historical figure. Stein himself makes it clear that he has aimed to understand what the historical figures thought. He says of his theses regarding mechanism, for instance, that "they concern the program in the sense of what was actually done and what was actually intended, for the most part, by its principal exponents" (Stein, 1958, 172). And one can obviously fail in so distinguishing their own view from their subject's. Failure to do this, Stein suggests, is the reason for Koyré's faulty comparison of Galileo and Descartes. While Stein considered his interpretation of Galileo through the lens of Archimidean Platonism "interesting" and "essentially correct", "his treatment of the historical material, in the light of his interpretive principle [Archimidean Platonism], appears to fall short, to be in a certain sense unphilosophic" (Stein, 1958, 82):

> What Koyré does can be characterized in Platonic terms: he contrives a likely story, but fails of the truth because he has eschewed the more arduous task of a genuine confrontation of reality. That just this is a, perhaps the, characteristic pitfall of philosophy—one might therefore say, "the unphilosophical fallacy" par excellence—is one of the principal theses of the present study; and if the analysis contained in the foregoing excursus is sound, it may serve as a minor example to show that that fallacy, if absolutely inescapable, can at least be relatively avoided.

The distinction, then, means here to rule out "the unphilosophical fallacy" of confusing one's view and interpretive principles with the view of the figure in question.

However, this understanding may still masks two related errors. Recall the distinction between a theory's looking good and being true, introduced in §3.2. There, it was emphasized that the air of certainty around 'being true' was merely apparent, and this was demonstrated by introducing as equivalent the "economic" manner of speaking. Here, the air of certainty surrounds 'the philosophy of science *per se*'. Making the first error, one might take their own philosophy of science to be the ultimate one (Stein, 1958, 172-3):

> There is always a tendency to equate the doctrines of "this our knowing age" with that myth (cf. Huygens's references, on successive pages and clearly intended as equivalent, to "the Philosophy of the present day" and "the true Philosophy"), but this is not a philosophical tendency even though not unknown among philosophers.

The obvious way in which the certainty needs to be stripped is to recognize that we have hardly more right to claim ours is "the true Philosophy" than did Huygens his own. Thus he continues,

> To make the distinction cogent, *per se* must really be attached to philosophy-of-science rather than to science: we must distinguish between what certain men—contemporary or not, scientists or not—have thought about science, and what we think about it.

The second error is less obvious. Just before his remark on Huygens, Stein tells us that "it is well to bear in mind, in making such a distinction, that "science *per se*" is a kind of myth, the science with which we are in actual fact acquainted being unavoidably historical." One way in which the science with which we are acquainted is unavoidably historical is just that we write of what has already been produced. But what follows suggests he means something more radical:

> But what we think about it [science] is directed to some (historical) "it," and if we think of science as not only a historical product but an enterprise, then what scientists thought about it—what they thought they were doing when they did it—is a significant part of "it" itself.

What appears essential is thinking of science as not only a historical product but an *enterprise*. Insofar as it is an enterprise the role of its agents must be recognized; hence, *our* philosophy of science is informed by what was *theirs*. What this means is that we need *not only* to recognize that we have no claim to the true philosophy, but that, moreover, *because* we have no such claim, post-hoc reckonings with scientific theories *must* attend to what those involved thought. Thus, I suggest that avoidance of the "unphilosophical fallacy"—the "characteristic pitfall of philosophy"—in this context would mean also recognizing science's nature as a flesh-and-blood enterprise, both figuratively and literally.[7]

Committing this fallacy seems to support three distortions that will be relevant in the next section. Though perhaps not so distinct, I will speak of them as if they are for convenience. One such tendency, prevalent with respect to mechanism as elsewhere, has been to view theories as in a sense reified—that is, as (synchronically) stable and epistemically homogeneous collections of principles, results and/or representations regarding a fixed domain of phenomena. However, recognizing the significance of agents when analyzing the history undermines this, at least so far as mechanism is concerned. This program existed in layers, Stein emphasizes (Stein, 1958, 187-8):

> About the higher layers—about the general scheme involving the conceptions of mass and force, of momentum, the conservation of momentum, etc., and indicating the mode in which one ought to attempt to analyze any given natural phenomenon—there was by the last quarter of the seventeenth century a very large measure of agreement... But on the deeper levels there was not only very much less agreement, there was also very much more explicit hesitation, and toying with alternatives, on the part of individual thinkers. *The mechanistic program was not, even as a program, a finished conception.* It pointed, for the investigation of nature, in a certain direction; but it did not completely define a route. It evoked—somewhat variously for investigators imbued with varying philosophic predilections or intellectual dispositions—a certain image of the structure of nature; but it did not clearly exhibit what ought to be the elements of a complete picture of the world.

[7] A note on Stein's uses of 'philosoph-': At least in his dissertation, and I submit it as representative, there appear to be two primary uses. One—often 'philosophy'—tends to be used as a shorthand for the professionalized discipline, and hints of irony are sometimes detected. The second—often an adjectival/adverbial variant, e.g. 'philosophical'—is commonly used in a more genuine classical sense, so that 'philosophical' means something like 'comporting with genuine concern for the growth of knowledge and understanding.'

As a matter of practical consequence, this means that such reification and emphasis of theories papers over not only variation in epistemic commitment but also *what* is being committed to. Instead, Stein takes the view "of the "ideas" of science (i.e., the most general *systematizing* principles) as constituting a program rather than a doctrine" (Stein, 1958, 284); in brief, these programs offer—and are seen by its best prosecutors as offering—*regulative* rather than *constitutive* guidance (Stein, 1958, 283).

Another common (and related) distortion concerns the "dynamics" of theory evolution. A famous example, brought to us by Kuhn, is to view the major such changes as wholesale conceptual replacement. This view of theoretical development is, at the very least, not warranted with respect to the mechanistic program. A view of science as a thoroughly progressive development of theories is likewise inadequate. The dynamics is much more complex than either of these views (Stein, 1958, 251):

> ... *if* the development of modern physics from the sixteenth century is viewed as a continuing search for the really basic character of natural processes, then *it is possible to see, behind all the revolutions that the subject has undergone, a very direct continuity both of aim and of method*; continuity, indeed, in all senses of an ambiguous phrase, in "what the science has been about."...
>
> It is the fashion in these days to refrain from seeing history as a uniform progress toward the one, the true, and the good; and this fashion seems on the whole amply justified. But when the history of physics is viewed in the way we are considering, an exception has to be made of it: there is no denying, as a sheer empirical fact, that the attempt to comprehend phenomena by the Newtonian mechanical program and its successors has been attended with overwhelming, progressive success.

Such subtleties are present also in that "paradigmatic" example of a Kuhnian revolution, namely, the developments at the foundations of physics at the turn of the twentieth century. In that episode there was a kind of continuity as well, preserved through the reinterpretations and redefinitions of central aspects of the mathematical tools at play (Stein, 1958, Ch. VII).

This understanding of program "dynamics" hints at a necessary expansion of what theories can be. The question Stein asks himself with respect to the "dynamical" aspect of theories is not whether, e.g., mechanism is being treated as a theory, for "in one sense every program implies a theory; or at least a hypothesis, namely the hypothesis that the program can be successful" (Stein, 1958, 284). Rather,

> The appropriate question, then, is not whether mechanism has in any sense been treated as a theory in this discussion, but whether it has been treated as a theory in an illegitimate sense, that is whether we have supposed susceptible of empirical confirmation a doctrine that is not in fact capable of being either established or overthrown by experience.

The answer is no, for such "doctrine" *is* capable of being either established or overthrown by experience (Stein, 1958, 288). Thus, views of the "higher layers" of programs as *mere* conventions are, if taken too seriously, in danger of misrepresenting the enterprise of science. We will see this mistake soon in Poincaré's philosophy of science.

Whereas the first two distortions involve a kind of reification and homogenization of theories, the third seems to involve not only this but also a homogenization of their relations to the world, as well as a subsequent reification of, as it were, the theory's image of the world. Of course, so far as mechanism is concerned, there is the problem, mentioned above, that it was not a finished conception and therefore did not have a settled "image" of the world. But putting this aside for the moment, Stein, I suggest, is still troubled by such a "crisp" view of theory-world relations. This appears implicit in his discussion of the following characteristic of the mechanistic program (Stein, 1958, 169):

> The test of experience in the narrower sense, the test of any particular explanation (theory) by confrontation with phenomena, was taken in a very stringent way. Not only the general aspects, or certain particular details of the phenomena, but (in principle) *all* details were expected to be rendered adequately by the theory.

The point I'm concerned with comes in his clarification of the reservation implied by 'in principle'. This clarification comes in three "heads", and the second is most relevant here. It begins (Stein, 1958, 170):

> *Second*, the theoretical conceptions that are brought to bear in the mechanical explanation of phenomena are not only abstract, in practice their application always involves some oversimplification; and this means that the agreement to be expected between observation and theory is limited by some reasonable estimate of the errors entailed by that simplification. The oversimplifications stem from two causes: lack of detailed information about the actual conditions, and the exigencies of calculation. For example, in discussing the motions of the planets from the viewpoint of the theory of gravitation it is simplest to treat the planets as masses each of which is located at a single point. To a second approximation, the shape of the planets may be taken into account by regarding them as spheres. . . Next, by taking into account the oblateness of the planets, in particular of the earth, one can deduce a further motion that is in fact observed in the precession of the equinoxes.

What I see as latent in this observation is that such knowing oversimplifications complicate the ways theories "represent" the world, their diversity implying, in turn, a diversity of ways of "representing." This is clear in realizing that our "reasonable estimate of the errors entailed" by an oversimplification, which inform our expectations for agreement between observation and theory, is itself informed by our understanding of the role that simplification plays in the situation. For this reason, assuming theories represent the world in largely homogenous ways—or, in more explicitly enterprise-y terms, that scientists *believe* that their theories are this way; believe that they come with something like a basic, general set of rules for how to apply them[8]—is not appropriate.

Thus in this way, too, assuming a theory implies a crisp world-according-to-theory "image" is unwarranted. What this means in practice is that a scientist's "belief" in a theory needn't engender strict reductionist dismissal of other researches by dint of their not conforming with its "ontology". And at least so far as mechanism is concerned, it didn't. That is, (Stein, 1958, 172)

> the "reduction" of phenomena to "mechanical causes," as that envisaged in the mechanistic program, does not entail the concentration of scientific interest upon phenomena of motion, to the neglect of the intrinsic qualities and patterns of things. Fresnel's mechanical theory of light, for example, is not primarily a theory of the jostlings and tuggings of particles of ether, but of such qualitatively optical phenomena as the formation of images, double refraction, patterns of color and shade produced in diffraction, and so forth.

So however it was that mechanists envisioned the world through the eyes of the program, it was not so psychologically domineering as to suppress more "ontologically" nebulous inquiries.

[8]"In short," Stein says, "[Newton's program] is not a program that works automatically" (Stein, 1958, 96-7). See Chapter V, especially pages 96-100.

1.3 "Yes, but…": a dialectical undoing of the RID

In the last section I argued for three main points. First (§2.1), I argued that Stein is consciously wary of the distortions and dangers of undue generalizations, certainty and clarity. To establish this, I showed that the opening remarks of "Yes, but..." allude to and thematically mirror those that framed his dissertation, wherein this fact is apparent. Second (§2.2), I argued that Stein disapproves of the rhetorical ends to which such undue generalizations, certainty and clarity have been put in inquiry itself. Ultimately, this came to his appreciation that science is a dialectic, hence his turning away from the anxiety for ultimate clarity and certainty and his championing of Newton's genuine skepticism in the study of natural phenomena. Third (§2.3), I argued that Stein disapproves of the ways such undue generalizations, certainty and clarity have been generated from the history. This was evinced by his reckoning with the subtle and pervasive consequences of "the unphilosophical fallacy," or what is the same, his affording of respect to the many layers of complexities of the flesh-and-blood enterprise of science. I will now bring these observations to bear on "Yes, but…".

The structure of Stein's argument in "Yes, but…" can be confusing read linearly. For this reason, what follows is not a line-by-line or paragraph-by-paragraph reading. Instead, in light of the §2 discussion, I will present it as a kind of "dialectical undoing": I will describe Stein's reckonings with the presuppositions and misconceptions that underpin the RID in the order of (my estimation of) their importance. My aim in this is to better emphasize how the dialectical conception of science and the unphilosophical fallacy guide the three major motifs of the argument in "Yes, but…". Each of these will be given their own subsection. The concluding subsection will consist of a brief characterization of what Stein takes to follow from the three major motifs.

1.3.1 The dialectic of science

The first motif of the argument in "Yes, but…" actually pushes us toward Stein's own dialectical view. This view is already doing heavy lifting by the second page. By then, he has already ruled out trite instrumentalism and expanded the scope of a sopisticated instrumentalism to include all of "the world of experience." This immediately rules out any (even epistemically "principled") instrumentalism that tries to delineate what can be (or is) known of the world from what cannot be (or is not) (Stein, 1989, 56). Thus, in Kantian terms, we are already empirical realists. Like in his discussion of Mach, we are ruling out any laying of constitutive principles and thereby rejecting the faulty-empiricist double standard. Stein therefore assumes that the instrumentalist's claim can be read as: "A theory is "nothing but" an instrument for representing phenomena" (Stein, 1989, 50).

Stein characterizes this as a "somewhat liberalized instrumentalism." Whether or not Stein really thought this was only *somewhat* liberalized, this is clearly not true: what Stein in fact describes is a *maximally* liberalized instrumentalism, being "instrumentalism" only in the sense that it is presented in the economic *manner of speaking*. When Stein later elaborates on the first irony of the RID's treatment of this history by considering the case of Poincaré, this is clear (Stein, 1989, 56):

> The contested reality was of the ether; and Poincaré, because he regarded the ether as a fiction rather than a reality, was unwilling to take very seriously (although he was willing to play with) the idea that charged particles exchange momentum with the ether. This, however, is a very odd position for an instrumentalist to take (the first irony); for there is no warrant at all in the instrumentalist view for *grading* the entities of a theory in *degrees* of reality or fictitiousness—regarding particles as *more real* than the ether.

Like the energeticists and Mach, Poincaré had "subtly transformed" the economic manner of speaking into what I've called the economic *view*. He did this by speciously distinguishing between charged particles and the ether, despite their being equally real by his own standard (Stein, b, 21-2); Poincaré had effectively endorsed the double standard Stein sees as recurring in the empiricist tradition (e.g., in Locke (Stein, 1990b, 33-4), Hume (Stein, 1990a, 209;219)(Stein, 1993, 189-90), Mach (above)).[9]

The sophisticated "instrumentalism" in play throughout "Yes, but..." is therefore only a different manner of speaking. All it serves to do is emphasize that our knowledge and understanding are still subject to expansion and revision. So when Stein remarks that the empiricist double standard is "a bit of unregenerate realism, doing the work of the Devil among the empiricists and instrumentalists" (Stein, 1989, 56), he is not alluding to the realist and instrumentalist doctrines familiar from the RID; rather, he is saying the following: the misguided empiricist has (i) succumbed to the tendency to speak in a realist manner in the cases that are "closer to home" and an economic one elsewhere (perhaps because of some anxiety for ultimate clarity and certainty), but (ii) failed to recognize that these are, at bottom, *just* manners of speaking, and therefore (iii) mistaken the linguistic shift for something real. This he sums up in closing (Stein, 1989, 64-5):

> What we are left with is that other, provisionally, "ultimate" or unexplained fact, that we *do* find ourselves compelled to formulate our beliefs in non-phenomenalistic terms; and in this process, atoms, electrons, fields, and the like are in a case quite analogous to that of chairs, tables, and the like in Berkeley. The *justifiable* claim to "reality" possessed by [sic] those "theoretical entities" is of the same kind as the justifiable claim—not after all denied by Berkeley—of these ordinary objects.

[9]Luckily Einstein, unlike Poincaré, took the theory seriously as a guide to how the world is and explored its implications (Stein, b, 23). Lorentz, too, took the theory seriously (Stein, 1987, 389-90) by recognizing the same distinction as did Maxwell between "what is known with some security, or held at least with some probability, and what is bare and even implausible conjecture" (Stein, 1989, 62).

To hold this is to reject the faulty-empiricist double standard; and "realism" in this sense I endorse unreservedly.

1.3.2 "Nothing but" and "something more"

The second motif concerns the instrumentalist's claim that theories are "nothing but" instruments and the realist's that theories are "moreover" true. However, because Stein has rejected the faulty-empiricist double standard, the "instrumentalism" of which he speaks is rather *anti-realism*. This is to say that Stein's dialectic view suggests that the RID is, in its purest form, the *transcendental* debate between realism and anti-realism (e.g., idealism).[10] The question being debated is thus whether the noumena are anything like the phenomena. Berkeley, for one, was an anti-realist. His doctrine was "that what is real is just minds and their perceptions, and that *all* our beliefs about the physical world are just "instruments" for organizing and anticipating experience" (Stein, 1989, 61). Stein asks of this doctrine: can it be refuted? *Need* it be refuted?

One might think that, since a phenomenalistic basis does not appear possible, Berkeley is after all wrong: it cannot be that all of our beliefs—especially our belief *in* such a world—are but instruments. But this does not get at (what Stein believes is) Berkeley's essential point, which is that there is an insuperable gap between the world and our experience of it. As we may view Kant as doing, the failure of phenomenalism could just be considered a fact about us rather than the world as it is independently of us. How could you possibly respond to an anti-realist like Berkeley? If even the failure of phenomenalism is irrelevant, it seems nothing we know—in the everyday sense of the word—would satisfy the anti-realist. Need one, then, reply to the anti-realist? No, Stein says—they are "irrelevant to any real issue in

[10] It is seemingly for this reason that when Stein later refers to this paper, he calls it "an article devoted entirely to the issue of realism and anti-realism" (Stein, b, 22).

the understanding of science" (Stein, 1989, 65); there is no need to reply to the anti-realism of, e.g., Berkeley, Kant, or later Putnam.

Some have risen to the anti-realist's challenge, however. This is the case, for instance, with Boyd's proposal for determining "real" validity over "mere" instrumental validity—the latter meaning that a theory affords a correct and adequate representation of phenomena. Boyd and his anti-realist interlocutor have both assumed this instrumental validity from the outset Boyd (1983)[11]. The question Boyd then tackles is thus understood as: how do we determine whether theories are "moreover" true *of the noumenal realm*? Boyd claims that evidence that theories are "moreoever" true comes from considering "the connection of theories with the *ongoing process of scientific inquiry*" (Stein, 1989, 51). However, because Boyd's claim is *at best* an hypothesis—it in principle could not have been "induced by the [noumena]," for we have no access to that realm—it cannot explain; his purported explanation is "*disconnected from its explanandum*" (see: color location problem, Huygens (Stein, 1989, 53-5)). This is in contrast to Huygens's hypothesis regarding the constitution of light: *that* hypothesis led rather directly to further investigations of optical phenomena, and in this sense was a good hypothesis according to Stein. For Boyd's claim to likewise be a good hypothesis, it would need to lead to unmediated investigations of the *noumenal* realm as it really is. However, this is not possible *by definition*, so it can lead to no further inquiry. This is Stein's point (a): "argument to a better, or the best, explanation is a doubtful business, which I should prefer to view as abductive, or heuristic, or tentative at best; and in the present case, I do not see what investigations are to follow the abduction" (Stein, 1989, 52). Understood in this way, as a debate about transcendental semantics, Boyd's claim—like the anti-realist's—is irrelevant to understanding science.

But the situation is even worse than this, for the terms of the debate seem to rule out the identification of *any* non-transcendental "something more". This Stein summarizes in his

[11] Boyd's publication concurrent with "Yes, but…", Boyd (1989), also fits the description here.

point (b), that he does not believe that the explanation does in fact explain (Stein, 1989, 52-3). For suppose that Boyd's hypothesis does actually explain something about scientific theorizing and is thus evidence for the realist thesis. What is this thesis? It is that theories that are correct and adequate representations of phenomena "moreover" possess some extra attribute; i.e., we're putting aside those attributes had by correct and adequate theories, each of which presumably concern the theories, the phenomena, or relations thereof. By (a) we know that whatever this extra attribute is, it cannot concern relations to the noumena. But then this extra attribute must somehow concern either the phenomena, the theory, or relations thereof. However, whatever it is, this attribute would fail to be "something more"—it would be precisely one of those attributes put aside in the search for something more. Therefore, because the realist thesis fails to identify any (non-transcendental) extra attribute which is *to be* explained, we surely cannot identify evidence that *does* explain it. So either, as in the last paragraph, the realist thesis takes the noumenal sense, which means that it cannot be explained and is irrelevant to science; or we assume that Boyd's hypothesis explains, in which case it can't be explaining the realist's "moreover" thesis. Should the realist forego this "something more," they are left with no way to distinguish themselves from the dialectical and enterprising inquirer, whose explanation of theorizing they insist is inadequate.

It is in this temptation to respond to the anti-realist that the taste for genuine explanations is lost, and the arguments slide into, e.g., claiming that things would be miraculous otherwise. But an enthusiasm to find out need not, and in especially the best scientists *does* not, give way to such an anxiety in the face of "inexplicable" facts (Stein, 1989, 64):

> Indeed, if one examines the explanation offered by science today for the existence and properties of aluminum. . . , it is hard not to feel that this explanation, at least as much as Berkeley's, grounds ordinary things upon a miracle; but the simple fact is that whatever our science adopts, perhaps provisionally, as its "ultimate"

> principles, just because they have no further ground, remain “inexplicable”; and the farther they are from the familiar, the more they will seem “miraculous”.

In this sense, miracles are a natural part of the dialectic of science.

This loss of the taste for genuine explanation is made palatable by the Quinean motif of the “ontology” of theories (Stein, 1989, 57). This is because the motif does little to dispel the misconceptions noted in §3.3. Here it is relevant that the focus on “ontology” has tended to bring with it a view wherein theories induce (by a privileged semantics) a “crisply” reified image of the world. This leads naturally to a comparison problem: is the actual world—independent of our theorizing, and as represented with this privileged semantics—(approximately) the same as the theory’s image of it? The realists and anti-realists then look to the history to support their contrasting claims.

But this is to commit “the unphilosophical fallacy” in the several ways of the last section. Among the misconceptions that seem to get this comparison problem off the ground is the assumption that theories always imply precise characterizations of the world. This is evident, for instance, in the debate of Putnam (1975) and Laudan (1981) regarding the reference of ‘atom’ and ‘ether’. Each is assuming that a “proper” analysis of the theories in question will reveal that ‘atom’ and ‘ether’ each correspond to precisely-characterized aspects of reality that we now know to, respectively, be and not be present. However, this assumption is problematic. For one, each aspect in fact *does* appear present in substantial ways and, moreover, according to the same metric of “presentness” (i.e., being weighable). More generally though, our understanding of what is “implied” by the theories has been subtlized, demonstrating that even “fixed” theories don’t imply fixed, sharp images of the world.

This points to a broader sense in which the unphilosophical fallacy has been committed here. What underlies the assumption that theories imply precise characterizations of the world is the privileging of a robust semantics of theory-world relations. The central notion for Putnam and Laudan, for instance, is reference. However, the notion does not play for them the same role it does in the Tarskian semantics from which it appears borrowed.[12] Consider the term 'ether' as it occurred in the nineteenth- and twentieth-century theories (for the moment, let's assume these "theories" make distinct existence claims about the ether). Recall the T-schema: *'… ' is true iff … **, where ' '…' ' is a metalanguage name for a claim in the object language and '…*' is the metalinguistic translation of that claim. Let's consider, as a claim made by the nineteenth-century theory, the statement 'Light travels through an ether'. The T-sentence for this claim *as made in the context of the nineteenth-century theory* is then:

> 'Light travels through an ether' is true iff light travels through an ether*

We are presuming that this statement was taken as true then, hence the latter is satisfied. Then, the T-sentence for the claim *as made in the context of the twentieth-century theory* is:

> 'Light travels through an ether' is true iff light travels through an ether**

We are presuming that the claim is false, hence the latter is not satisfied. However, to say with Putnam and Laudan that the former is true while the latter is false, what has *actually* changed is the metalinguistic translation (noted with '**'), and therefore the base clauses in the metalanguage's definition of truth. This is precisely to note that 'light travels through an ether' is *not* a statement that is true or false *simpliciter*. Baked into this change is a fluidity of the notion of reference at play: the notion of reference one distills changes along with the change of metalanguage, and for this reason it takes a backseat to translation in Tarski.[13] In this sense reference and truth are trivialized in Tarskian semantics: they are (sometimes

[12] See Maddy (2007), especially Part II, for a similar and more careful treatment of truth and reference in scientific inquiry. Like Maddy, Stein appears to consider Tarski a disquotationalist (Stein, 1989, 50).

[13] For instance, the new metalanguage need not even translate the statement compositionally or in the same compositional manner, the former of which simply eliminates any meaningful role for reference.

useful) linguistic devices meant to capture obvious and basic features of linguistic use. The machinery doesn't tell us anything substantial about either how we do or ought to determine whether a statement is satisfied or how we do or ought to perform translations. What it *actually* seems to do is draw our focus to the changes in knowledge and understanding that inform these decisions.

There is essentially only one way for Putnam and Laudan to "untrivialize" reference and truth, though it comes in different forms. One form fixes the notion of reference and insists that translation and truth must run through it. This, in turn, means that there is an objective and precise fact about the referential relations enjoyed by a statement. One can then say that 'Light travels through an ether' was *considered* true by nineteenth-century theorists but was, in fact, *objectively* false; since the statement's referential relations do not change, there is no meaningful change of the metalanguage. A second form insists that there is an objective fact regarding the preservation of meaning by translation. What this boils down to is assuming there is one *meta*-metalanguage in which all equivalences across metalanguages can be judged: i.e., the meta-metalanguage allows us to form statements like *[Light travels through an ether*]*$^{+}$ *iff [Light travels through an ether**]*$^{+}$ that express the equivalence of metalinguistic translations of an object-language statement. This essentially allows us to express in the meta-metalanguage that the nineteenth-century theorist's metalanguage is not one that appropriately translates their object-language statement, so that they misunderstood the genuine meaning of their statement and thereby considered it *true* when it was, in fact, *false*.

These moves are essentially the same, however, in that each treats the semantics as beyond reproach, i.e. outside the dialectic. In the former form, the nineteenth-century theorist either held the wrong metalanguage or held the right one and misunderstood it; in the latter form, they either held the wrong meta-metalanguage or held the right one and misunderstood it.[14]

[14]The errors could, of course, have been in still-higher languages, but I ignore these cases because the main point is unchanged.

But in either case, it is assumed that we've reached the ultimate, not-subject-to-revision language in which we can evaluate "objectively."[15] Though it manifests differently, the effect is of a kind with the faulty-empiricist double standard: an end is put to inquiry. Insofar as this is considered a philosophy, it is also to commit the unphilosophical fallacy in the same way as Huygens: assuming one's own is the True philosophy, not subject to modification and thus treated as outside of the dialectic of science. The problem is thus not reference *per se* but assuming that there is a once-and-for-all way to capture word-world relations.[16] Because such assumptions lie "outside" the dialectic, this is to go "external" in the Carnapian sense.[17]

1.3.3 Excessive simplicity

The third motif is an attempt to take the debate as an honest empirical inquiry into the historical use of theories, despite this not being the aim of Boyd or Laudan. However, the outcome is still negative. Even taken in dialectical stride, reference as a robust concept is clearly too simplistic to represent theory-world relations. What these relations actually look like is better captured by what Stein calls 'theoretical structure'. However, one should take care not to read too much into 'theoretical structure'; what is meant is loose, something along the lines of 'conceptual system'. This is intimated in his nodding to Hilbert et al. (see Stein (1988)). For them, the goal was *Tieferlegung der Fundamente*, which I prefer to translate

[15]There is a certain methodological irony here, too: in assuming a True philosophy so as to provide an objective description of past theorizing, one plays down what can *more* fairly be called objective—descriptions of what the past scientists *thought* they were doing. Presumably, only the former description is open to revision via scientific developments after these past scientists' time.

[16]The assumption can be especially pernicious, as it has been in discussions of color as a Lockean quality: what was once taken to be a clear and justly-answerable question about the nature of things—what quality of bodies does color resemble?—should, by dint of subsequent research, be seen as fundamentally confused (Stein, 1989, 53-4). (In fact, research done *prior* to Locke's publication of the *Essay* showed this Stein (2004).)

[17]Recall that Stein has a quite liberalized notion of linguistic framework in mind. He is also acutely aware of the "mutual dependence of frameworks and theories," and therefore the danger that "the internal/external distinction may lead to the neglect of important large questions that span the development of theories" (Stein, 1970, 285-7). Thus, the internal/external distinction here is just between dialectical and non-dialectical inquiry. This use of the distinction appears to be one of the "valuable philosophical lessons" of logical empiricism, alongside the illuminating discussions of its flaws.

as ‘foundational deepen*ing*’ to more strongly emphasize its ongoing nature. Even at their most pristine, which is to say as a system of axioms, theories were constantly subject to modification, axioms being added, subtracted, modified, or even removed to elevate “mere” theorems. This was done according as our understanding of the world and the conceptual system itself changed. Hilbert, for one, considered this foundational fluidity in his axiomatic approach “a tremendous advantage”, for it makes room for conceptual and interpretational play while still allowing one to preserve what is considered essential mathematical/conceptual components in the process (for instance, see Frege’s excerpt of Hilbert’s 29.12.1899 (Frege et al., 1980, 41)).[18] Indeed, in pointing to quantum mechanics as one of his concrete examples, the ever-changing and idiosyncratic nature of how theories represent is a prominent part of Stein’s conclusions (Stein, 1989, 59):

> I do not claim to have a definitive formulation of the meta-physics of quantum mechanics; but I believe rather strongly that the difficulties it presents arise from the fact that *the mode in which this theory “represents” phenomena* is a radically novel one. In other words, I think the live problems concern the relation of the Forms—indeed, if you like, of the Instrument—to phenomena...

In addition to its suggesting that quantum mechanics represents phenomena in a novel way, this remark suggests that theories prior to it had “represented” in different modes, too. In equating ‘Forms’ with ‘Instrument’—which we know to be the acceptable economic manner of speaking of theories—it is underscored that Stein’s notion of structure is intentionally nonspecific, riding on his nonspecific characterization of theories.

[18] To get a better feel for what I mean by conceptual/interpretational play see Weatherall (2018), where it seems even *profoundly* formal approaches to understanding theories (in the context of equivalence, in particular) inevitably involve extra-formal judgments of theory-world relations; the formalism doesn’t just “spit out” an interpretation. This should be still more obvious with theories axiomatized *à la* Hilbert or with those not presently susceptible to even this approach, the latter of which appear to be most common.

Such uses of reference, as in the case of atoms and the ether, moreover reveal a misconception of theories as stable and homogeneous sets of beliefs such that we can specify their every representational feature. They are believed stable in the sense that what is and is not "in" the theory is clear. This is leveraged for claims like '*nineteenth-century electromagnetism* implied the existence of an ether' to appear sensible, when it is, in fact, often difficult to draw the line between a theory and its interpretation.[19] They are moreover assumed homogeneous in that their hypotheses and demonstrations are treated as on an epistemological par. This gives rise to statements like '*mechanism* assumed all aspects of the world are explainable ultimately in terms of particles'. But this assumes both an agreement on high-level mechanistic conceptions as well as a strong reductionism which did not exist generally among its practitioners.

Correcting these misconceptions of theories as stable and homogeneous also corrects assumptions about how theories evolve and what they are for. When it is claimed, for instance, that 'ether' was believed to refer but was later discovered not to, this is taken to reveal a fairly discrete change in theoretical commitments. Subtleties of what was actually shown aside, this appears to assume that theoretical advancement is straightforwardly represented with something like a network of propositions: nodes as beliefs, edges as inferential connections, and when those beliefs that "impinge" on experience are overturned, we strive to "mutilate" as little of the network as possible in accommodating the change. But this assumes a lot. Not least, it assumes that all the important features of scientific inquiry are appropriately captured by looking at *propositional* and fundamentally *representational* claims about how the world is. However, important features arise also from intimations concerning aspects of this knowledge itself, including not only the pedigree of the individual nodes but of the connections among them. In at least this sense, science is not only an enterprise of *knowledge* but also of *understanding* Stein (2004). Thus, when Boyd calls attention to "the connec-

[19] Thus it was that Hertz's famous remark that "the Maxwell theory is the system of Maxwell's equations" was actually a reflection of a deep scientific achievement. See Stein (1970, 281-2).

tions of theories with the *ongoing process of scientific inquiry*" (Stein, 1989, 51), Stein sees this as a call to attend to these extra- or merely heuristically-representational aspects of theories—what may broadly be called methodological, as opposed to metaphysical aspects.

Theories are therefore *for* more than representing, and as a consequence their merely representational aspects must be considered in that broader context. In particular, it would appear ill-advised to play one off the other in the way some RID arguments have tended to. Laudan, on one hand, focuses on the representational aspect of theories, taking their historical alternations to tell us that theories are "mere" instruments; Boyd, on the other, appears to focus predominately on their methodological aspect, taking the historical continuity to tell us that theories are "moreover" true. The RID thereby tends to generate from the history "conclusions whose contraries may equally be "demonstrated"" (Stein, 1958, 9) because of an inadequate understanding of the principles guiding the debate. Both the "wanton" and "teleological" interpretations of the history have *some* truth to them; however, as a point of historical fact, neither adequately characterizes any substantial swathe of inquiry on its own. Lost in the back-and-forth are the varied ends to which theories have been put *by science's agents*: Resources for inquiry—Yes, but. . . representational—Yes, also.

1.3.4 No difference that *makes* a difference

Finally, where does the argument of "Yes, but. . . " leave us with regards to an understanding of scientific inquiry? The emphasis just placed on science's agents in understanding theories suggests one direction: taking seriously what they think about science and what they think they're doing when they do it. In so doing, it also suggests still more concerning the "wanton" and "teleological" views of inquiry. Not only are they and their presumed opposition—as well as their hidden philosophical and historiographical premises—inadequate as a matter of historical fact, they are moreover of dubious worth even as guides to further inquiry.

We recognize this, Stein's work seems to suggest, when our focus is returned to inquirers themselves. This focus is evident in what is arguably *the* central thread of Stein's philosophy: if and how individual inquirers can claim to know anything when they can believe nothing certainly.

As I hope to have shown in §2, Stein was already heartily pulling this thread in his dissertation. There, the thread culminated in the following question (Stein, 1958, 392):

> Is the degree of dispassionateness in ultimate formulation, which in the view here presented philosophy seems to require, compatible with the degree of passionate conviction required for a scientist like Newton, or Maxwell, or Einstein, to devote himself to the development of his "daring hypotheses" in the teeth even of great opposition?

As we've come to expect, he answers this question by appealing to an example he considers representative. Funny enough, the example will be familiar to readers of "Yes, but..."—for it is precisely the one he gives to clarify and make plausible his "no difference that *makes* a difference" claim: Maxwell's systematic removal of his ether model from the theory of electromagnetism. Unlike in "Yes, but...", however, he there makes explicit what answer this example is to give. What this answer is should, by now, come as no surprise: "Here the history of science is rather encouraging to the old view of man as capable of rationality" (Stein, 1958, 392-3).

I wish to recommend we take seriously Stein's belief *that* it is possible to reconcile these two aspects of genuine skepticism as a call for inquiry into *how* this is so; I suspect this is what he intended, too.[20] In taking this inquiry seriously, I think it is appropriate to see this as a *development* of the RID into a *related* question that is both relevant to scientific inquiry (and

[20]For one, Nancy Nersessian's doctoral dissertation was titled *Scientific Evolutions: On Changing Conceptual Structures in Science*, and her research since has taken a robustly interdisciplinary approach to this cluster of issues. She completed her dissertation with Stein (Nersessian, 1977).

thereby, philosophy of science) and more ripe for attack than it was 30 years ago.[21]. This is no small undertaking; it is likely to span not only the specific sciences themselves, but much of the cognitive and psychological sciences, such as perception and attention, concept formation and structure, learning and expertise, motivation, biases and constructivity, the structure and functions of memory, developmental, social, and educational psychology, and aspects of psychology still unconceived or unrecognized.[22] Naturally involved will be a careful attention to how various aspects of cognition, including mathematics, are *embodied*, have evolved, and are culturally shaped. To put things in Stein's own words, the broader question is this (Stein, 1989, 55-6):

> . . . how [is it] that our own natural endowment—which has evolved for its "instrumental" value in coping with far more immediate aspects of the world—has also proved to be an "instrument" capable, under favorable circumstances, of (e.g.) discovering quantum mechanics[?]

1.4 Conclusion

In closing, I would like to return to that fated remark—that what Stein really believes "is that between a cogent and enlightened "realism" and a sophisticated "instrumentalism" there is no significant difference—no difference that *makes* a difference" (Stein, 1989, 61). On its face, as Stanford intimates, the claim appears to be about two views *per se*: upon finessing,

[21] I am borrowing from Stein the phrase and conception of question-changing in the growth of understanding and knowledge (Stein, 2004, 165).

[22] These issues are highly complex and interrelated, too, necessitating extreme care when considering "big pictures" of the role our endowment plays in how inquiry works. To give one example demonstrating (the *possibility* of!) a kind of upward percolation of conceptual revision in cognitive research: recent results (Winter et al., 2016) from the centroid paradigm (Sun et al., 2016) suggest the potential incompleteness of the feature integration theory (Treisman and Gelade, 1980), the latter of which influences our understanding of the binding problem as well as Getalt principles. Changes in our understanding of the latter could rather quickly force modifications to any "big picture" of the nature of knowledge (as evidenced in, e.g., (Spelke et al., 1998)).

the views *themselves* converge such that they are no longer distinct. Understood this way, it would seem that the RID is still "well-joined" in the sense that its broadest ambitions are clear and well-founded but that (i) the realist and instrumentalist views so far proffered in the RID are wrong, and (ii) the correct view is something of a blend of the two.

This isn't quite right. As the scare quotes seem to indicate, we aren't dealing with *views* at all, let alone views proffered in the context of the RID. On the one hand, a sophisticated "instrumentalism" recognizes that theories are *for* representing in addition to being resources or "instruments" for inquiry; it has been "sophisticated" in that it has traded in the anti-realist's philosophical skepticism for the genuine skeptic's reserve. Having done so, the error of the faulty empiricist double-standard is apparent, and thus returns the capacity for *genuine* discrimination by fidelity to detail. On the other hand, a cogent and enlightened "realism" recognizes that theories are resources or "instruments" for inquiry in addition to being *for* representing; it is "cogent" and "enlightened" in that it has traded in the transcendental philosopher's pseudo-answers—explanations disconnected from their explanandum—for the genuine skeptic's discrimination by fidelity to detail. Having done so, the error of the faulty realist's claim to the True philosophy is apparent, and thus returns a *genuine* reserve. With the hubris of the RID realist and the despair of the RID instrumentalist removed, there is no clear debate left: the genuine skeptic makes quick dialectical work of the oversimplifications born in the realist and instrumentalist of the "unphilosophical fallacy", and with these goes, too, the judiciousness of realism *versus* instrumentalism.

What is therefore left are two *manners of speaking*: a "realism" whereby one speaks passionately of what the world is and isn't like, so far as we know, and an "instrumentalism" whereby one speaks dispassionately of the forms of our current theories while still taking them seriously as about the world. To speak in either manner is to strive to live in accordance with the twin virtues of genuine skepticism—to practice balance and judgment. That these different manners of speaking—passionately with Maxwell as a "realist" or dispassionately with Lorentz as an "instrumentalist"—*make* no difference in the practice of our best scientists, as the history suggests, is then an empirical fact about us in our relation to the world. So for at least as long as it remains cogent to say *that* these manners of speaking differ while making no difference, we should seek to understand *how* this is so.

Chapter 2

History Can't Save Aimless Philosophy of Science

> Those who think that history repeats itself—whether as tragedy or farce—are condemned to misunderstand it. (Rowe, 2018, x)

2.1 Introduction

Since Kuhn, the pendulum has swung heavily toward history in philosophy of science inquiries. It is undoubtedly a good thing that philosophers have expanded their attention to include science's history. Even so, we do not often discuss how we make use of science's history in our philosophizing. One broad-strokes method for integrating history in philosophizing is what I call *History Implies Norms* (henceforth HIN). As the name suggests, philosophers using this method aim to generate norms for contemporary science from its history. This method has been used, for instance, in the realism debate as of late. However, it is unclear whether the turn to history in philosophy of science has led to norms any more useful than before. Here I present a pessimistic response. I argue that the debates we have, and methods we use therein, are unable to deliver the normative guidance that many philosophers of science seek.

My argument proceeds as follows. First (§1), I introduce the HIN method for integrating history into philosophy of science, and I present several prominent attempts to use this method. This culminates with Stanford's HIN argument for being an instrumentalist, which I take to be the most serious (and prominent[1]) HIN argument available today. Then (§2), I evaluate Stanford's argument as the exemplar of general philosophy of science's use of the HIN method. After showing that he must provide a causal argument for his central claim, *Historicism*, I argue that his argument presents evidence having neither the pattern of an inductive argument (§2.1) nor of a counterfactual analysis (§2.2). Given that one of these patterns must be used to establish *Historicism*, and that *Historicism* is his only argument for being an instrumentalist, he has therefore given no argument at all for being an instrumentalist. Finally (§3), I argue that not having concrete goals is responsible for philosophy of science's failure to derive useful norms. After showing that Stanford faces a dilemma between backing off his rhetoric or improving his methods (§3.1), I argue that he faces the dilemma because the scientific realism debate has no concrete goal (§3.2). Thus, I conclude with the provocative claim that introducing history into our normative philosophy has made no difference because philosophers of science generally lack concrete goals, hence philosophy of science lacks a serious methodology.

[1]Indeed, Synthese has dedicated an entire special issue to Stanford's Problem of Unconceived Alternatives (Bhakthavatsalam and Kidd, 2019).

2.2 Connecting History and Normative Philosophy of Science: History Implies Norms

In philosophy of science using the HIN method, the history of scientific methods and their outcomes constitutes a body of evidence that conditions outcome expectations for future uses of those methods. Despite being a commonplace way of thinking, it arguably reaches its zenith in science itself. Suppose I need to get from the second floor balcony of my office to the first. I can either jump or take the stairs. Historically, in similar situations, jumping has led to injury and taking the stairs not. I wish to remain uninjured, hence I expect taking the stairs to be the optimal choice. In this way, rudimentary observations from the past, which we have organized into theories of kinesiology and mechanics, have conditioned my expectations for the future. That is, they inform what I *should* and *should not* do. HIN functions likewise. I begin by identifying my goals and, at least crudely, the acceptable options for achieving them. Then I look to similar situations in science's history to evaluate these options. These past situations condition my expectations for the consequences of my current choice. Thus, given their particular goals, the history of science tells contemporary scientists what they *should* and *should not* do—science's history *implies norms* when conjoined with scientists' goals. Henceforth I speak of *history* implying norms for simplicity's sake, suppressing the necessary reference to goals.

Kuhn is easily read as using the HIN method. Against the backdrop of Popperianism, where history was used not as evidence but merely as (purported) illustration of the norms he specified, Kuhn meant to effect a "historiographic revolution" (Kuhn, 1970, 3). This historiographic revolution arose from the failure of the cumulative image of science as it confronted science's history. The revolution made new theories of science possible, ones constructed more directly from science's history. Kuhn's own theory undoubtedly conditioned our expectations for the future of science, and Kuhn hence implied norms. And, indeed, this was intended (Kuhn, 1970, 207–8; emphasis added):

> A few readers of my original text have noticed that I repeatedly pass back and forth between the descriptive and the normative modes, a transition particularly marked in occasional passages that open with, "But that is not what scientists do," and close by claiming that scientists ought not do so. Some critics claim that I am confusing description with prescription, violating the time-honored philosophical theorem: 'Is' cannot imply 'ought.'
>
> That theorem has, in practice, become a tag, and it is no longer everywhere honored. A number of contemporary philosophers have discovered important contexts in which the normative and the descriptive are inextricably mixed. 'Is' and 'ought' are by no means always so separate as they have seemed. But no recourse to the subtleties of contemporary linguistic philosophy is needed to unravel what has seemed confused about this aspect of my position. The preceding pages present a viewpoint or theory about the nature of science, and, *like other philosophies of science, the theory has consequences for the way in which scientists should behave if their enterprise is to succeed.*

In brief, what scientist's *have* done tells us what they *should*. This is because science has made progress,[2] at least so far as individual scientists have achieved some of their goals, and obviously what scientists have done has clearly contributed to this. Thus he continues,

> Though it need not be right, any more than any other theory, it provides a legitimate basis for reiterated 'oughts' and 'shoulds.' Conversely, one set of reasons for

[2] Kuhn was clear that, whatever our measure of progress, science has made it: "Though scientific development may resemble that in other fields more closely than has often been supposed, it is also strikingly different. To say, for example, that the sciences, at least after a certain point in their development, progress in a way that other fields do not, *cannot have been all wrong*, whatever progress itself may be. One of the objects of the book was to examine such differences and begin accounting for them" (Kuhn, 1970, 209; emphasis added).

> taking the theory seriously is that scientists, whose methods have been developed and selected for their success, do in fact behave as the theory says they should. My descriptive generalizations are evidence for the theory precisely because they can be derived from it, whereas on other views of the nature of science they constitute anomalous behavior.

However, Kuhn's use of the HIN method was not entirely successful. In particular, readers are hard-pressed to draw crisp 'shoulds' from Kuhn's theory. Take, for instance, Kuhn's account of scientific change. This account distinguishes between two kinds of science: normal and revolutionary. Crudely put, normal science operates under a host of shared commitments; it is conservative by nature. Revolutionary science is instead characterized by crisis, i.e., a fracturing of shared commitments in the face of troublesome anomalies; it is inventive by nature. This account clearly differs from the Popperian one. On that account, revolutionary science is "like normal science but better" (Bird, 2018). Yet the import of Kuhn's new account for scientific practice is unclear. For a start, should we pursue more revolutionary science by more routinely denying shared commitments? Kuhn seemingly thought not. But common invocations of Kuhn's account suggest, yes, we should more routinely deny shared commitments. So even assuming there is one, the correct answer is apparently not obvious. Consequently, practicing scientists are unable to draw clear norms from Kuhn's account of scientific change, for surely not all scientists "behave as the theory says they should," meaning the norm for individual scientists cannot be as simple as "keep doing what you're doing."

Laudan's use of the HIN method in the so-called *pessimistic induction* (PI) was similarly unsuccessful. This induction famously culminated in Laudan's list of "successful but false" theories (Laudan, 1981, 33). Since these theories were successful but false, we likewise cannot infer truth from the success of our current theories. More circumspectly, Laudan summarizes his means–ends proposal as (Laudan, 1978, 140–1)

> the sketch of a *naturalistic* theory of methodology which preserves an important critical and prescriptive role for the philosopher of science, and which promises to enable us to choose between rival methodologies and epistemologies of science. What it does *not* promise is any a priori or incorrigible demonstrations of methodology; to the contrary, it makes methodology every bit as precarious epistemically as science itself. But that is just to say that our knowledge about how to conduct inquiry hangs on the same thread from which dangle our best guesses about how the world is.

Laudan is clearly using the HIN method: the history he tells is supposed to imply norms. But as with Kuhn, it is unclear what the norms are. Consider the following. The PI is supposed to push us to think of science as problem solving. This means that evaluating a theory fundamentally involves comparing it to its rivals. With what voracity, then, should a scientist pursue a rival? In particular, should they be more or less voracious than before? The answer is unclear. So again, whatever norms may be implied by Laudan's history are insufficiently precise to guide action.

Stanford's use of the HIN method makes similar normative promises. Undoubtedly, Kuhn's and Laudan's accounts were suggestive. However, they failed to identify a descriptive difference among scientists that makes a normative difference *to* them.[3] Stanford thinks he has identified one. The difference is whether one believes in the existence of *unconceived alternatives* (UAs). That is, have we failed to conceive of "fundamentally distinct alternatives to extant theories that [are] nonetheless both scientifically serious and reasonably well-confirmed by the evidence available"(Stanford, 2019)? If you believe UAs exist, then you are a *historicist* or *instrumentalist*; if you do not, you are a *realist*. Stanford has long

[3] Although Stanford does not use the "difference that makes a difference" language, borrowed from Howard Stein, until later, this is nevertheless the spirit of his program as early as Stanford (2001). Note, too, that, whatever the form of this "difference," it must manifest in different behavior by scientists. Science, after all, is practiced!

claimed that this difference of belief will make a difference in our science (Stanford, 2006, 211):

> ...the realist and instrumentalist differ perhaps most radically and most fundamentally in their conceptions of and expectations concerning the future of scientific inquiry itself. The realist supposes that our further inquiry into the natural world will continue to bear out at least in broad strokes the various conceptions of domains of nature that are articulated by our current scientific theories. But the instrumentalist offers us a very different picture of the future of human scientific inquiry. She judges it quite likely that even the most genuinely impressive and instrumentally accomplished theories of contemporary science will ultimately be replaced by more powerful conceptual tools offering fundamentally different conceptions of nature that have not yet even been conceived. She rejects the idea that even the most fundamental claims of theoretical science will persist indefinitely into the future as part of the best collection of conceptual tools we have for engaging the natural world. And among sensible people who are rightly impressed by the dramatic empirical success of the best scientific theories we have, what greater difference could there be?[4]

Indeed, this is a constant in Stanford's writing after *Exceeding Our Grasp*.

However, Stanford appears to deliver on his promise to derive norms from science's history, unlike Kuhn and Laudan. In Stanford (2016, 332–333), he echoes the above practical differ-

[4]The passage that culminates in this paragraph highlights a substantial tension in Stanford's book. On the one hand, there is supposed to be a meaningful difference between how realists vs. instrumentalists do their science. I take it that this difference manifests in their inferences. Yet on the other hand, this "does not mean the moral suggested here is that we must somehow constrain and regulate the inferences we draw in the course of our scientific theorizing by some perfectly abstract and general commitment to the likely existence of completely unspecified but serious unconceived theoretical alternatives"(Stanford, 2006). Indeed, Darwin, Galton, and Weismann each wrongly believed the conclusion of their eliminative inferences, but they nonetheless remain blameless. I see no way to dissolve this tension, so I will henceforth assume that Darwin et al. were blameworthy.

ence between realist and instrumentalist science. But he also spells out what these differences are:

> Moreover these differences in turn produce a further and equally profound divergence concerning the actual pursuit of scientific inquiry itself: a scientific instrumentalist of this historically motivated variety will be systematically more sanguine than her realist counterpart concerning the investment of time, attention, energy, taxpayer dollars, and other limited resources in attempts to discover and develop theoretical alternatives that diverge in fundamental ways from or even directly contradict the most powerful and impressive scientific theory we have in a given natural domain.

So this difference *matters* insofar as Stanfordian instrumentalism tells us to dedicate more of our limited resources to discovering and developing theoretical alternatives. Of course, the realist will have motivations to seek these alternatives, too. However, the instrumentalist has "at least one more that is far more compelling: in stark contrast to the realist, she fully expects this search to ultimately attain its intended object" (Stanford, 2016, 333).

Therefore, Stanford appears to have successfully used the HIN method: from historical premises, in conjunction with an assumed goal, he derives normative conclusions. Stanford states this argument in numerous places, most recently in Stanford (2020, 10–13). However, its clearest enunciation is the following (Stanford, 2019, 3931; emphasis added):

> The question, then, is not whether revolutionary or transformative science is a worthy goal, but instead whether we expect to reap sufficient benefits from more aggressively pursuing it to outweigh the inevitable costs of doing so. Elsewhere I've argued that *the answer we should give to this question depends in important ways on the position we take in the ongoing debate concerning scientific realism*

> *itself*: after all, the need for "revolutionizing entire disciplines; creating entirely new fields; or disrupting accepted theories and perspectives" [cit. Bement] is considerably less pressing if scientific realists are right and their historicist opponents are wrong than if the reverse is true (Stanford, 2015).

Working backward from the problem he identifies, the argument is this. In contemporary science there exist UAs that we could conceive of if we dedicated more resources to searching for them. Call this kind of UA *imminent*. We have not dedicated more resources to searches because we are too conservative with funding. In turn, we are too conservative with funding because we are realists, i.e., we do not believe such searches will succeed. Thus, if we were instrumentalists instead, we would fund more liberally; and if we were to fund more liberally, then we would conceive more of the imminent UAs. In short, Stanford's positive claim is

Historicism: If you are an instrumentalist, then you will conceive more of the imminent UAs.

Before proceeding, I want to emphasize two points about *Historicism*. First, it is crucial here that *Historicism* only considers *imminent* UAs. Were it to concern any UA, it would imply, for instance, that Newton was a realist because he failed to conceive of relativity theory. This is plainly absurd: substantial mathematical, conceptual, and evidential barriers lay between Newton and relativity theory, so failing to conceive of relativity theory should not condemn him. Instrumentalism simply could not have advanced Newton's science by two centuries. Thus, *Historicism* must concern only those UAs a scientist could have reasonably conceived of in a given situation. Obviously this requires many aspects of the scientist and their world to align just right, including sociopolitics, psychology, and even health, in addition to the obvious scientific aspects. Regardless of what this alignment consists in for a given scientist and unconceived alternative, if it does so align I call the UA *imminent*. Stanford is not so careful as to speak only of imminent UAs in his work; in fact he freely uses relativity as an example of a UA for Newton alongside what appear to be examples of imminent UAs (e.g., (Stanford, 2017, 217)). (I suspect this "lack of care" owes to the move discussed in fn. 6). Nevertheless, I make this distinction on his behalf and assume it has been satisfactorily defined. The definition's details do not matter for present purposes.

Second, I have chosen the form of *Historicism* above from among two possibilities. I could have chosen the inverse, i.e., *In-Historicism*: if you are a realist, then you will conceive fewer of the imminent UAs. Stanford does appear to endorse *In-Historicism* at times, too, particularly in more recent work (e.g., (Stanford, 2019)). All the same, I do not think this choice matters because an argument analogous to the one that follows in §2 can be given for *In-Historicism*. Moreover, the burden is on Stanford to disambiguate the two claims, and that he has not further supports my contention in §3 that philosophers of science generally do not appreciate the tension between their methods and their goals. That said, I have chosen *Historicism* here because the work I discuss, *Exceeding Our Grasp*, emphasizes individuals as the level of analysis. Concerning individuals, I expect Stanford wants to claim not that realism "always harms" but that instrumentalism "can't hurt" progress, whereas I expect he wants to claim the reverse in his later population-level analyses.

Therefore, if we assume with Stanford that science should pursue any and all imminent UAs,[5] then the norm is clear: be an instrumentalist. In what follows, I will use '*Instrumentalism*' to refer to this norm.

[5]This is implied when Stanford assumes that successfully discovering and developing theoretical alternatives is essential to discovering and developing "even more instrumentally more powerful successors to our best scientific theories," where the latter is assumed to be the goal of scientific activity (Stanford, 2015, 877). I won't address this except to say that (i) Stanford does not argue for these assumptions and (ii) neither assumption is value-free. I have argued for the latter elsewhere.

2.3 *Historicism* Requires a Causal Argument

I claim that Stanford has provided no argument for *Historicism*. The reason is elementary: the argument Stanford provides has the wrong logical form.[6] To see this, note that *Historicism* is a causal claim: if you adopt an instrumentalist rather than realist attitude, then this causes you to conceive more of the imminent UAs. Thus, his argument must establish a causal link between instrumentalism and conceiving more of the imminent UAs.

Roughly speaking, there are two distinct kinds of argument he could give to establish this causal link.[7] I address them separately in §2.1 and §2.2. Regardless of the argument's form, the argument, if there is one, is contained in *Exceeding Our Grasp* (henceforth *EOG*). We gather this from the fact that, in his relevant later works, he argues only that science has become more theoretically conservative. Whenever Stanford moreover claims that instrumentalism is the solution, the argument of *EOG* is implicated.[8] Hence, without any apparent loss of generality, I will only discuss the argument as it appears there.

2.3.1 Argument for *Historicism* by Induction?

Consider Stanford's argument for *Historicism* as an induction. Stanford describes *EOG*'s argument this way in the book itself (Stanford, 2006, 19). This description invites a statistical analysis. Since *Historicism* is a causal claim, merely correlational evidence is unable

[6]This situation is complicated by Stanford's move to the contrast between Catastrophism/Uniformitarianism. In short, this move "subsumed his position [here understood as a "ground-level" critique of eliminative inferences] in a more generic historicist critique" (Maddy, 2020, 35), along with the PI. Yet what recommended Stanford's use of the HIN method over Laudan's was the specificity of *Historicism*—i.e., it told us what to do. The generic historicist critique does not. Therefore, I ignore this move in what follows.

[7]To be clear, I claim that Stanford *must* provide an argument of one of these kinds, not that he *intended* to do so.

[8]For example, Stanford (2019) claims that instrumentalism is the solution to theoretical conservatism and refers to Stanford (2015) as evidence that instrumentalism would make this difference. The latter then refers to *EOG* as evidence (i.e., an argument for *Historicism*) for *Instrumentalism*.

to establish it. Rather, the argument must rely on experimental evidence.[9] If Stanford's argument provides experimental evidence, we should expect its model to look like the following. First, the independent variable (IV) is the presence of realism or of instrumentalism for an individual with respect to a theory, and the dependent variable (DV) is the presence or absence of imminent UAs in their relevant subsequent work. Second, we identify our sample population: scientists whose theory possessed an imminently conceivable theoretical alternative. Third, we assign instrumentalists to the experimental group and realists to the control group. We conclude by identifying the null hypothesis, that there exists no effect on the DV by the IV.[10] I now argue that Stanford's actual argument is not experimental because it looks nothing like this model.

An immediate concern is how representative Stanford's sample is of the general scientific population. His argument consists of investigations of Charles Darwin, Francis Galton, and August Weismann. Several concerns have been raised about Stanford's sample. For one, it has the bad company of the PI since the latter, too, only draws on science from before the 20th Century. Both Park (2011) and Fahrbach (2011) argue that any sample is unrepresentative if it does not include theories post-1900.[11] Lewis (2001), and then Magnus and Callender (2004), each argue that the PI embodies the base-rate fallacy. Mizrahi (2013) likewise criticizes the generality of the PI. An inadequate sample appears blameworthy for these criticisms, too. For another, Stanford's sample itself has come in for criticism. Mizrahi

[9]Strictly speaking, the argument can at best be *quasi*-experimental, given that we cannot manipulate the independent variable.

[10]I have simplified and supplemented Stanford's argument in numerous places here. Here are the most salient. First, I have assumed that he is concerned with individual scientists' attitudes and subsequent science, rather than a community of scientists. While later work casts doubt on this assumption, *EOG*'s focus is clearly on individuals. The assumption is significant as it informs the logical form I've given *Historicism*, i.e., whether instrumentalism (realism) is necessary or sufficient for conceiving of imminent UAs. Nevertheless, the burden is on Stanford to specify this; moreover, this further supports my overall point that philosophers of science using the HIN method do not appreciate the tension between their methods and their goals. Second, I have modelled the IV and DV as dichotomous. However, there is no loss of generality, provided the relationship between the variables is expected to be approximately monotonic, and I expect Stanford is committed to this.

[11]Park argues that the "explosion" of scientific knowledge after 1900 means a fair sample would not only include theories of the 20th century but moreover consist primarily of these theories, since they so vastly outnumber previous theories.

(2015), for instance, argues that Stanford's sample is both too small and too ambiguous to decide between a pessimistic or optimistic view of the truth of our present theories.

I am sympathetic to criticisms of Stanford's sample, but the model above makes a still larger problem apparent. Let us consider the control and experimental groups in turn. Recall that the control group is made up of realists. *Historicism* implies that much, perhaps most, of the science produced by this group will not contain imminent UAs. That is, the realists will fail to consider the imminently conceivable theoretical alternative in their work. It is uncontroversial to say that some realists have failed to conceive of imminently conceivable theoretical alternatives. However, the precise distribution of DV values matters for rejecting the null hypothesis and, hence, establishing causality. Yet determining this distribution does not matter for us. Stanford's argument has a much more fundamental problem when viewed as experimental.

The most significant problem is that Stanford constructs no experimental group. Recall that the experimental group is to be made up of instrumentalists. *Historicism* implies that only some, perhaps few, of these instrumentalists will fail to conceive of imminent UAs in their work. We may grant this as plausible. Yet, there is simply no evidence to support this claim. Indeed, while Stanford gestures throughout his publications at further control group cases, I have found zero experimental group cases. Without an experimental group, there is no way to even *test* the null hypothesis, let alone reject it. Experiments are fundamentally comparative, and yet no comparison has been made. At the very least, then, Stanford must fill out the experimental group.

Nevertheless, let us imagine some experimental group cases. On the one hand, there are many cases where it appears that instrumentalism coincides with conceiving more of the imminent UAs. Consider the history of quantum mechanics, for instance, where interpretations appear to have proliferated in the absence of realism about the Copenhagen interpretation. Some have even argued that, in my terms, these alternatives were imminently conceivable at the time the Copenhagen interpretation coalesced (e.g., (Becker, 2018), (Cushing, 1994)).[12] Nevertheless, it remains to be seen whether one or more of these alternatives gains traction. If not, then Stanford cannot rely on his typical tool—coming to later acceptance—to argue for their UA bona fides. On the other hand, there are plenty of cases where instrumentalists failed to conceive of serious alternatives. Newton, for instance, was famously non-committal as concerned the efficient cause of gravity.[13] Yet, we had to wait for Einstein to get relativity theory. This is, of course, a silly example, and it illustrates the reason for requiring UAs be *imminent*. On yet another hand, we have the strange case of Poincaré who, in a certain sense, authored a theory he himself failed to conceive of (Stein, c); it is not entirely wrong to say that Poincaré's instrumentalism is to blame. None of these kinds of cases support *Historicism*: the QM alternatives may not be UAs, relativity was not imminent for Newton, and instrumentalism is partly responsible for Poincaré's theoretical stasis. Undoubtedly, there are cases that support *Historicism*, too. However, we must see many more cases before we can be confident that the latter dominate.

But realists and instrumentalists should not be too hasty in fleshing out this model. As I elaborate in §3, the larger problem is not that HIN philosophy of science is using the right methods in the wrong ways; rather, it has no serious methodology at all, in large part because its goals are not sufficiently specific to constrain it. Unsurprisingly, then, pursuing this model would bring further challenges. I note two. First, the empirical validity of the concepts involved here is unclear. We do not know, for example, if the (informal) measures Stanford uses for realism and instrumentalism are construct-valid, i.e., that they measure

[12] I have argued elsewhere that such claims are dubious (Mitsch, 2020a).

[13] This is a long-running interpretation of Newton; see, e.g., Reid's 16th December, 1780, letter to Lord Kames (Reid, 1852, 56–60). Seemingly of the same mind as Stanford, there Reid writes that "'never to trust to hypotheses and conjectures about the works of God, and being persuaded that they are more like to be false than true' [quoting his Lordship]" "...is the very key to natural philosophy, and the touchstone by which everything that is legitimate and solid in that science, is to be distinguished from what is spurious and hollow." Note that I do not interpret "hypothesis" to be synonymous with "explanation by efficient cause;" this means Newton could have legitimately (by his own lights) desired an efficient-cause explanation for gravity without thereby "feigning hypotheses" (see, e.g., Chen (2020) on Newton concerning his attitude toward efficient causes of gravity). This is non-trivial, of course, but that actually speaks to my overall point: substantial interpretive work is needed before *Historicism* can be established.

what they say they measure. One can read Mizrahi (2015) as an argument that more work is needed on this point. Similarly, we should expect these concepts to be value-laden, and this must be addressed. A comparison with a similar case of philosophy–psychology crossover seems to bear this out (Mitsch, 2020b). Second, the model and study design deserve further work. Sampling from history is difficult, and new methods, like a model-based approach to sample representation (see, e.g., (Zhao, 2020)) or those used in bibliometric work (see, e.g., (Yan et al., 2020)), may be valuable. Still, perhaps these significant challenges are surmountable and *Historicism* may yet triumph. Regardless, the argument as it stands has not established causality by induction. It would need to be experimental, which requires having an experimental group. Plainly, it does not have one.

2.3.2 Argument for *Historicism* by Case Study?

Now consider Stanford's argument for *Historicism* as consisting of case studies. This means that the argument deals with instantiations of *Historicism* of the form "If SCIENTIST had been an instrumentalist about THEORY, then they would have conceived of ALTERNATIVE." However, *Historicism* is a causal claim, and purely descriptive case studies cannot establish causality. This is because claims like the previous instantiation are counterfactual claims, and describing merely what actually happened does not tell us what would have happened under some historical alteration. This presents a serious problem for HIN philosophy of science. However, this is a problem outside of philosophy of science, too. In particular, policymakers often must rely on studies of as few as one case to determine what to do. Determining what to do involves distinguishing causal connections, hence methods have evolved for doing this. The method used in political science is *Comparative Counterfactual Analysis with Process Tracing* (CCAPT), which I discuss here for its simplicity and whose uses I should not be understood as universally endorsing.[14] In this section I introduce CCAPT,

[14]There is a sense in which CCAPT is the only option available since any alternative will look quite similar. Thus, CCAPT can be likened to an informal presentation of Causal Bayes nets (see, e.g., (Glymour et al.,

and then contrast Stanford's argument for *Historicism* with a recent argument that explicitly uses CCAPT. In doing so, Stanford's argument for *Historicism* appears plainly inadequate.

CCAPT is a combination of techniques used to avoid errors in historical counterfactuals. The gist of CCAPT can be seen if I introduce the argument I contrast with Stanford's, which I have chosen for its popularity: Harvey (2011)'s argument concerning the cause of the Iraq War. There, Harvey addresses the prevailing narrative that "[i]n essence, neoconservatives, backed by other senior members of the Bush administration, abused their control of the White House to push the country into a war of choice that would otherwise never have happened" (Harvey, 2011, 1). He calls this account of the Iraq War *neoconism*. This account thus holds that, counterfactually, a Gore administration would not have gone to war. Call this counterfactual *Gore-peace*. Because this counterfactual grounds the causal claim of *neoconism*, it must be argued for. Nevertheless, Harvey had found no argument for it in the literature. That is, no counterfactual analysis had been provided. Moreover, the alternative—that a Gore administration would also have gone to war; call this counterfactual *Gore-war*—had not been considered. But if *Gore-war* is true, then *neoconism* is false. Thus, Harvey performs a comparative counterfactual analysis (CCA) of *Gore-war* vs. *Gore-peace* to determine the strength of the *neocon* account of the Iraq War. Naturally, fleshing out each counterfactual involves "[tracing] empirically the temporal and causal sequence of events within a case that intervene between independent variables and observed outcomes" (Bennett and George, 2001). That is, it involves process tracing.

CCAPT is a natural method to use for HIN philosophers of science, like Stanford, who typically rely on case studies. Let me use the case of Darwin and Pangenesis as a stand-in for each of Stanford's cases. *Historicism* implies

2010)). The major conclusions that follow would remain unchanged if translated into the latter framework. Also, compare what follows to Bolinska and Martin (2020).

Instance: If Darwin had been an instrumentalist about Pangenesis, then he would have conceived of explanation by common cause.

Note that I have substituted 'he would have conceived of explanation by common cause' for 'you will conceive more of the imminent UAs.'[15] This is appropriate because, as Stanford himself seems to gesture toward (Stanford, 2006, 69–71), explanation by common cause was imminently conceivable by Darwin. In this way, its conception is our proxy for theoretical conservatism. So *Instance* for Stanford is as *Gore-peace* is for *neoconists*. Analogously, corresponding to *Gore-war* is

Counterinstance: Even if Darwin had been an instrumentalist about Pangenesis, he still would not have conceived of explanation by common cause.

The remainder of the section will contrast Harvey's and Stanford's arguments according to the following CCAPT conditions, taken from Levy (2015): clarity and specificity; plausibility; cotenability; consistency; and comparative. Since Stanford's argument is not a conscious CCAPT, I make friendly modifications where possible. Throughout, I am supposing that Stanford's argument meets the obvious condition of historical accuracy.

Before moving on, it is worth noting that Stanford should believe two claims about Darwin. First, Darwin was a realist about Pangenesis. Second, this caused him to fail to conceive of explanation by common cause. This is because *Instance* can only be true if realism was responsible for Darwin's theoretical stasis; after all, Stanford offers instrumentalism as remedy for the effects of realism. I suggest we read him this way for the following reasons. Stanford certainly believes the first, as pages 65–8 of *EOG* make clear. There, he presents convincing evidence that Darwin had not only conceived of no alternative to Pangenesis,

[15]The details of Pangenesis and explanation by common cause do not much matter for my purposes, so I forego their description. All that matters is that explanation by common cause was, for Darwin, an imminent UA with respect to Pangenesis—or, so Stanford says.

but moreover that this fact grounded his claim that Pangenesis would reappear as the true explanation of heredity. This is just what it means to be a realist. It is less clear whether Stanford believes the second. In pages 70–5, Stanford suggests that common cause could have been accepted by Darwin and argues that Darwin failed to comprehend common cause when Galton proposed it to him. Stanford moreover suggests that Darwin would not have accepted common cause because it minimized the phenomena Darwin primarily meant to explain with Pangenesis. For present purposes let us suppose this implies that Darwin failed to conceive of common cause because of his realism about Pangenesis. Thus, we may reasonably attribute belief in *Instance* to Stanford.

Clarity and specificity

First, consider the clarity and specificity of the antecedent(s), consequent(s), causal path(s), and nature of the relationship between the antecedent(s) and consequent(s) involved in Harvey's counterfactual. The antecedent and consequent here are sufficiently clear and specific: Gore wins the presidency (rather than, e.g., "a Democrat wins the presidency"); the US invades Iraq (rather than, e.g., "the US coercively threatens Iraq w.r.t. WMD development"). Moreover, they are not *so* specific that their satisfaction becomes insufficiently plausible; for instance, it would be rather unhelpful to specify that "invades Iraq" comes down to precisely "begins bombing Iraq the *morning* of March 20, 2003" or "Gore wins Florida because Florida's then-Secretary of State Katherine Harris failed to account for her clock being 7 minutes behind, allowing Palm Beach County, submitting at 5:06:47PM on November 14, to officially meet the extended deadline .[16] The clarity and specificity of the causal paths and nature of the relationship between antecedent and consequent in Harvey's case are somewhat less obvious, in large part simply because it involves the machinations of a complex system

[16]Of course, one might still consider this information relevant for purposes other than the counterfactual being considered.

(international relations) over roughly a year and a half; however, suffice it to say that the 300+ page book at least addresses this concern.

The antecedent and causal path of Stanford's *Instance* appear less clear and specific, but not egregiously so. At this stage it seems reasonable to call the antecedent sufficiently clear and specific. After all, all Darwin needed to do was believe that there were serious alternatives to Pangenesis. Likewise for the consequent, Darwin merely would have needed to entertain common cause as an alternative. These *seem* straightforward enough to suppose Stanford escapes this comparison with Harvey relatively unscathed.

Nevertheless, the consequent of *Instance* presents a difficulty because the actual case of Darwin does not concern a UA. Strictly speaking, the alternative was not a UA because it was not unconceived; as Stanford explains, Galton had afforded him two opportunities to understand common cause. Darwin failed, of course, but presumably failing to understand is not the same as failing to conceive. So it seems that Stanford meant for misunderstood alternatives to also count as UAs. However, this only complicates the present situation. Part of what makes *Instance*, hence *Historicism*, feel so intuitive is that we have a fairly firm grip on what it means for a scientist to conceive of an idea. It is much less clear what it means for a scientist to understand an idea. Yet whatever it means, it further muddies the causal path. Now we must ask not only *why* Darwin did not understand Galton but *how* his understanding would have manifested. Indeed, had he understood and nevertheless rejected common cause, it is unclear whether UA instrumentalism remains the intuitive remedy rather than a nearby idea like intellectual humility. Inevitably, how one gets from the antecedent to the conclusion will be affected by what the conclusion is, so the specificity of the causal path is diminished by that of the conclusion. Thus, Harvey seems to have bettered Stanford with respect to clarity and specificity.

Plausibility

Second, consider the plausibility of the antecedent of the counterfactual, as well as the causal connections and interdependencies involved therein. The plausibility of these is directly related to the magnitude of their rewrite of the historical facts, i.e., the more plausible the smaller the rewrite. This stricture on plausibility is commonly called the minimal-rewrite condition. The condition is clearly related to the first criterion, of clarity and specificity. For instance, Gore winning the presidency is typically considered a minimal rewrite because it was eminently possible: either (Harvey, 2011, 25)

> a few more hanging chads during the recounting of precinct votes in Florida following the 2000 US presidential election; or an alternative US Supreme Court ruling on the methods used in Florida to recount electoral ballots; or Al Gore's successful bid to retain the electoral college votes by winning his own state of Tennessee; or a decision by Ralph Nader to pass on running in the 2000 election[,]

each could have led to a Gore presidency. Very little needs to differ from the real world to satisfy this counterfactual antecedent.[17] Likewise, given Gore's well-documented call for a stronger, more forceful multilateral foreign policy; the bipartisan pressure for the U.S. to secure Congressional or U.N. approval of a multilateral solution, rather than act unilaterally; Saddam's insulation from accurate information concerning the seriousness of the U.S.'s threat to invade; the dearth of good intelligence sources in Iraq after 1998; etc., the causal connections and interdependencies that obtained in the *actual* world seem appropriate for the counterfactual as well. Since not much changes in Harvey's counterfactual antecedent, it is at least plausible that not much needs to change to carry us to the consequent.

[17] Just to note a relevant connection here with broader historiography: if you're enamored of a "Great Men"—or its modern alternative, individual leaders—approach to history, the construction of *relevant* plausible counterfactual antecedents will often be easier. Structuralists will often have their work cut out for them because of the attendant complexity.

This minimal-rewrite condition has a corollary of direct relevance to Stanford's argument: generally speaking, counterfactual antecedents in which a historical actor is to undergo drastic cognitive changes are not minimal. This is because one would have to anticipate the changes that would *allow for* these cognitive changes. Were this to concern a capricious and willfully-naive individual—say, a Donald Trump—then the rewrite might still be minimal; a rewrite could be as simple as, say, Trump choosing to wear a mask or missing an episode of *Fox and Friends*. However, this is quite different from the more typical case of a fairly stable, well-intentioned and well-informed individual who, for subtle biographical and environmental reasons, fails to conceive of an alternative. I take it that this is the more typical case for scientists. Indeed, this may be another reason why Stanford is hesitant to blame Darwin, Galton or Weismann (Stanford, 2006, 135–6).

This manifests in Stanford's other cases, too, by the way. Let us consider "August Weismann's Theory of the Germ-Plasm" for a moment. This case makes clear a kind of "nested hierarchical structure" of theoretical stasis: because Weismann failed to conceive of an alternative at one upstream point in the theoretical structure of inferences, downstream inferences, too, are infected by theoretical stasis (Stanford, 2006, 128–9). What Stanford concludes is that even if Weismann had considered, somewhere along the line, an alternative relevant at that upstream point, it matters that he did not do so *at that precise point*. Thus, conception of alternatives is context-sensitive by being tied to a particular inference. Indeed, even if, contrary to fact, his peers had considered the alternative in some other context, the alternative remained unconceived in the UA sense. But this context-sensitivity illustrates another hurdle to fleshing out an alternative where Weismann's instrumentalism led to his conceiving the imminent UA. Now, a theorist must not only be an instrumentalist about some particular theoretical claim, they must be so in just such a way that it manifests in the precise context necessary for the conceived alternative to count *as conceived*. This is asking a lot. One of the major justifications for using counterfactual analysis is the fact that people often consider alternative futures when they are presented with a choice. When we construct a historical counterfactual in this setting, we are assured that the counterfactual's antecedent is plausible because the *historical actor* saw it as an alternative future. By definition, nothing so neat can happen in Stanford's cases, whereas one of Harvey's central points is that Gore's would-be administration would have seen an alternative future that led to war. Thus, it is unclear what evidence we can point to in justifying the antecedent's plausibility.

Notwithstanding the plausibility of the antecedent itself, we should consider the causal connections and interdependencies involved in its supposition. As I noted regarding clarity and specificity, the consequent, hence causal path, of *Instance* is less clear than it initially seemed. Recalling that Stanford's ultimate goal was for scientists not merely to conceive of but also develop theoretical alternatives, let us suppose the consequent now involves Darwin investing significant time and energy in the common cause alternative. If *Instance* is still to be true, this complexity must propagate backward. That is, either Darwin's instrumentalism or causal connections to it must be responsible for Darwin investing this time and energy. This demand is highly non-trivial, and we should expect many variables to be involved—many more, in fact, than for Harvey, who need only assume that Gore's assertive multilateralist foreign policy would have led him to support the policies he did in the actual world. To my mind, this places counterfactual arguments like Stanford's in something of a dilemma. Either select only obviously problematic examples (e.g., Trump), which will inevitably be of limited use in diagnosing broader problems in scientific practice; or, select only the ecologically typical cases (e.g., Darwin), which will inevitably involve counterfactuals of such complexity that little certainty or generality can be extracted. Needless to say, I think Harvey's counterfactual is more plausible than Stanford's.

Cotenability

Third, consider the cotenability of the antecedent with the additional assumptions needed to sustain it. As Levy intimates, much would need to be considered in Harvey's case: "[t]he election of Al Gore as US president in 2000 would have changed not only the belief systems and personality of the president but also the identity and views of the president's leading advisors and top-level government officials, the relationship between the president and the Congress, and possibly other variables as well" (Levy, 2015, 394). Indeed, in addition to these variables Harvey considers: domestic politics and the role the media and public might have played; the role of American intelligence; international politics; and Saddam's own failures and miscalculations, among others.

Now consider the cotenability of Stanford's antecedent with the additional assumptions needed to sustain it. Here we have a more basic problem: since we do not know what it would mean for Darwin to understand or what must have been different for this to come about, we cannot know what these additional assumptions are. This makes it impossible to consider their cotenability with his antecedent. However, let us consider some reasons why Darwin *might* have understood Galton's common-cause idea. Recall, as Stanford hastens to point out, that the idea Darwin struggled to understand in December of 1875 "was simply not a feature of the scientific landscape prior to Weismann's publication of it in 1883" (Stanford, 2006, 75). This seems to give us two routes to get Darwin to understand. First, Darwin might have come up with the idea on his own. In this case, we would also need to explain what Darwin understood of Galton's 1865 presentation of the idea (Stanford, 2006, 72). Perhaps he fully understood then. Yet then we must change Darwin's history before (at least) 1868. For example, maybe Darwin spent more time working through the text. But then, did he still publish his 1868 book? And we must also consider the ramifications after. For example, maybe he was so excited that he reached out to Galton before the latter's 1869 book. But if so, how would the readers of the 1869 book have been affected? On the other hand, if he did not understand in 1868, we must change Darwin's history between then and 1875. For example, maybe he began studying representative governments, giving him access to the analogy Galton used to distinguish his idea. But if so, why? Was there a civil uprising spurring his interest? And again, we must consider the ramifications of this change e.g., did he fail to publish anything further?... Or second, the idea may have entered the scientific landscape prior to Weismann's publication of it in 1883; this could be due to all sorts of social, political, or personal reasons, concerning any member of the scientific community (actual or counterfactual). Yet again, Stanford's argument appears worse than Harvey's.

Consistency

Fourth, consider the consistency of Harvey's causal story with well-accepted theory and empirical evidence. This desideratum is meant to restrict the causal path available in counterfactual constructions to only those that have significant empirical support. Put another way, it is meant to rule out the use of ad hoc theoretical arguments to sustain a counterfactual. However, as for philosophy of science, for international relations there aren't many well-established laws or theories; for this reason, Levy (and others) place emphasis on consistency with empirical evidence to so restrict the causal path. Harvey does this, to the extent that he does, by drawing on relevant theorizing/evidence from international relations, political science, media studies, security studies, and public polling, among other areas. For instance, in making the case that the media would have been just as complicit were Gore elected, he draws on a number of collections of domestic and international print and broadcast media taking the Bush administration's line at face value (e.g., Kagan, Enee, Moyer), as well as the lack of any substantiated criticism of that line (Harvey, 2011, 188–192). What this demonstrates, per Harvey, is *not* that the media was merely the mouthpiece of the administration, but that the global media had no administration-independent intelligence avenues.

Now consider the consistency of Stanford's counterfactual with well-accepted theory and empirical evidence. Here again we are in a bit of a bind. For one, the realism–instrumentalism debate has, historically, not led to much agreement.[18] Some entrants, in fact, have argued that the gap between the two poles should be made still wider (Park, 2019). Stanford himself has shifted the dividing line between the two poles on several occassions now. Thus, I am not optimistic that there is enough theoretical agreement among philosophers to significantly restrict Stanford's counterfactual. For another, we should want to ensure that this counterfactual about the psychology of Darwin is consistent with empirical evidence from the psychology and sociology of science. Yet this field is still in its infancy. Moreover, it is a massive enough undertaking just to get this evidence into conversation with the history and philosophy of science literatures,[19] let alone to measure the resulting body of work's consistency with Stanford's counterfactual. Harvey looks better with respect to consistency, too.

Comparative

Fifth, consider that Harvey's case is fundamentally comparative: *Gore-war* is evaluated right alongside *Gore-peace*. As Harvey emphasizes throughout his book, "evidence for and against both counterfactuals must be discussed together, because the strengths of one are directly relevant to uncovering the weaknesses of the other and vice versa. They should not be studied in isolation..." (Harvey, 2011, 27). Indeed, Harvey argues that counterfactual analysis is *inherently* comparative because of its logic: evidence for X's necessity for Y is evidence that X's absence is sufficient for the absence of Y. Thus, since it is inherently comparative, it is important that this be explicitly acknowledged in the structuring of one's argument so as to ensure genuine logical persuasiveness. Moreover, this is important not

[18]Stanford, of course, suggests otherwise, but that agreement seems to have been bought by Stanford at the cost of *Historicism*. See footnote 7.

[19]Though see (Nersessian, 2010) for an intriguing attempt to do just this.

only for logical reasons but also for psychological reasons. Such rigor seems well-advised in the face of mounting evidence that plot coherence differentially engages various large-scale neural networks, in particular modulating the recall of details (see Yang et al. (2019) for a recent meta-analysis; see Tylén et al. (2015) on the latter claim). In other words, failing to analyze comparatively can inflate estimations of plausibility.[20]

Now consider Stanford's counterfactual in its capacity as inherently comparative. Of course, Stanford *does* consider alternatives to instrumentalism—namely, variations on one very strong kind of realism (Saatsi, 2019), which he implies is responsible for theoretical stasis. This happens at significant remove from the particulars of the individual historical cases. Indeed, these *abstract* alternatives are what motivated *Historicism*. However, we've assumed that Stanford was arguing for *Historicism* by arguing for *Instance*. Yet to do this Stanford needs to expose *Counterinstance* as dubious. This is just to say that the abstract discussion should logically flow from discussions of individual cases. It does not, hence his argument is not comparative in the necessary sense for CCAPT.

Moreover, some of the evidence Stanford provides suggests that *Counterinstance* is no less plausible than *Instance*. As noted above, for a UA to be *conceived*, in Stanford's sense, requires that the UA is considered precisely in the context of the relevant eliminative inference. Thus for *Instance* to be true, Darwin must conceive of common cause in the context of inferring that Pangenesis is true. Suppose Darwin was an instrumentalist about Pangenesis. There are at least two reasons to suspect he would not have conceived of common cause. First, we know from Stanford that Galton twice presented Darwin with the common cause explanation. However, these presentations were given in the context of a downstream inference. Even after spending significant time and energy attempting to understand Galton's explanation, Darwin still failed to understand in this downstream context. Presumably, Darwin's path to conceiving of common cause in the context of the upstream inference goes

[20] Also see Farrell (2020) for discussion of this point as it concerns thought experiments in evolutionary psychology.

through understanding it in this downstream context. But then instrumentalism about the upstream inference is immaterial here: for one, the instrumentalism we've assumed must be localized to the upstream inference if it is not to be "pious and toothless" in the way fallibilism supposedly is (Stanford, 2006, 135); for another, Darwin *did* invest significant time and energy but still failed. It is hard to see how *Instance* is more plausible than *Counterinstance* when instrumentalism is immaterial at this crucial juncture for Darwin's conceiving of common cause.

Second, Stanford suggests that Darwin "almost certainly would not have" accepted Galton's alternative, even if he had understood it (Stanford, 2006, 75). It greatly matters here why Darwin would not have accepted common cause. Stanford suggests this is (Stanford, 2006, 75)

> in large part because he was increasingly convinced of the widespread existence and importance of the inheritance of acquired characters, whose significance Galton sought to minimize. Furthermore, Darwin was looking for a theory of inheritance that would permit natural selection to function as the engine of evolution, and he took this to require allowing an important role for the inheritance of acquired characters.

If this is correct, then Stanford should say that it was not realism about *Pangenesis* that caused him to fail to conceive of common cause, but rather it was realism about "the widespread existence and importance of the inheritance of acquired characters" as well as his own understanding of natural selection. An explanation like common cause simply was not what he was looking for. Of course, this reinforces my earlier remarks about the ambiguity and implausibility of *Instance*'s antecedent. Moreover, if the instrumentalism remains localized to Pangenesis *per se*, it speaks in favor of *Counterinstance* rather than *Instance*. Yet more generally, we should see the potential for paralysis when recommending nebulous changes in attitudes, like instrumentalism, versus specific behavioral or procedural interventions. Again, the instrumentalist attitude is put between a rock and a hard place: either it is "pious and toothless", like fallibilism, or it is particular to a cluster of inferences *whose makeup and logical connections we are generally unclear on.* In either case, specific guidance does not come easily. But providing specific guidance is the goal of HIN philosophy of science!

Stanford's Case Studies Do Not Support *Historicism*

Whether or not a convincing CCAPT argument for *Instance* can be given in the future, it is plain that Stanford cannot have done so yet. This is evident by contrast with Harvey's argument for *Gore-war* over *Gore-peace.* For one, Stanford does not adduce the evidence necessary to recommend *Instance* over *Counterinstance.* For another, even Harvey's argument has not been generally accepted[21]; this is to say, when they are done right, counterfactual analyses are incredibly demanding, and they still may not be accepted. Worse still, it is not even clear to what the truth of *Instance* would amount. Consequently, I suggest that Stanford's argument for *Instance* is not by CCAPT. We have then found that *Historicism* is supported neither by induction nor by CCAPT. Yet this leaves Stanford with no argument whatsoever, hence no reason to recommend instrumentalism.

2.4 What's the Point of Philosophy of Science?

The foregoing places Stanford in a dilemma. On the one hand, we might take Stanford's rhetoric at face value: he is asserting a norm (*Instrumentalism*) that is implied by a historical fact (*Historicism*). The foregoing shows that Stanford has failed in this case because his method cannot establish *Historicism.* On the other hand, we might take from Stanford

[21] For a brief introduction to its reception, I recommend Dawisha et al. (2013).

only what we can get: a response to interlocutors (scientific realists) in an abstract, normatively impracticable, debate (philosophy of science's debate over scientific realism). But then Stanford is no closer to genuine norms than were Kuhn or Laudan, and his "difference that makes a difference" language is inappropriate. In closing, I describe, diagnose, and suggest a treatment for this dilemma. The upshot is that HIN philosophy of science, like any other discipline, must reconcile its method with its rhetoric—that *history* serves as the evidence for HIN philosophy of science does not exempt it from elementary methodological rigor.

2.4.1 The dilemma: flawed argument or flawed rhetoric?

In §1 and §2 I took the former horn of the dilemma by taking Stanford's rhetoric—that he is asserting a norm—seriously. On this horn, Stanford gives an argument for adopting the norm that scientists be instrumentalists (*Instrumentalism*). Recall that this norm is recommended as the means for achieving science's purported goal of discovering and developing "even more instrumentally more powerful successors to our best scientific theories" (see fn. 5). *Historicism* is the inferential link that makes sense of *Instrumentalism* as a practical norm, since it essentially claims that scientists achieve their goal by following *Instrumentalism*. However, *Historicism* is then a causal claim about the effect of instrumentalism on conceiving alternative theories. As I argue in §2, Stanford gives a fundamentally flawed argument because he does not give a causal argument. Moreover, while I have not shown that *Historicism* is false, that even so brief a treatment as §2 belies *Historicism*'s intuitiveness suggests it is dubious. At any rate, Stanford has not yet established *Historicism*, and to do so would require a fundamentally different kind of argument.

Let me now take the latter horn of the dilemma by taking from Stanford only what his method shows. On this horn, Stanford has not established *Historicism*, hence he has not argued for *Instrumentalism* as a norm of science. This means that his rhetoric suggesting otherwise—that his argument is of the HIN form—is flawed. Instead, he has (at most) shown that scientists perennially fail to conceive of empirically well-confirmed fundamental alternatives to their current best theories, hence routinely err by using eliminative induction prematurely. In this way, he is engaging with opponents in the abstract scientific realism debate. Call this the *Problem of Unconceived Alternatives* (PUA).

The PUA does seem to have significant consequences for the scientific realism debate. Here are a few. First, the PUA can serve as a kind of "debunking" argument against realists who believe that the success of scientific theories would be a miracle if they weren't approximately true. The idea is that, whereas the "no miracles" realist makes the eliminative inference to theory realism because they can conceive of no alternative explanation for success, the instrumentalist familiar with the PUA knows that eliminative inferences of just this sort have routinely failed in the past. Second, the PUA demonstrates that scientists are bad at exhausting the space of theories that can account for a body of evidence. Third, the PUA gives the lie to any realist who confidently claims to know which part of a theory will persist indefinitely. Finally, the PUA points toward a non-realist conception of science by calling attention to science's perennial failure, like the PI before it. As with the first consequence, the last three also put realists in an awkward position.

Yet despite these contributions to the realism debate, the PUA is no closer to identifying a difference that makes a difference than were Kuhn or Laudan. Just as with the PI, the PUA does not tell us what to do regarding the pattern it identifies; we are no closer to a norm because we still cannot generate useful predictions about current or future theories. Indeed, there is a real sense in which *Historicism* has been *assumed for*, not *proven from*, the history Stanford provides. That is, his method only establishes the (purported) *problem* of UAs, not a solution, so if *Instrumentalism* is a norm that solves the problem, as he claims, Stanford must be assuming *Historicism* for his interpretation of the historical evidence for the PUA. In short, Stanford's rhetoric is flawed for papering over the fact that *Historicism* is assumed, not proven.

Thus, the dilemma: taking his rhetoric seriously, Stanford provides no evidence for a difference supposedly making a difference (realism vs. instrumentalism); or, taking his method seriously, he contributes evidence to a debate over a difference *that Stanford himself* claims makes no difference (the scientific realism debate prior to his entry). Either way, the history Stanford tells does not imply practicable norms. This contradicts Stanford's gesturing in (Stanford, 2019), where he implies that *Historicism* leads us to particular policy conclusions concerning funding and peer review. In fact, Stanford is in a worse position than his predecessors for making a significantly stronger claim on the basis of the same kind of evidence. As we saw in §1, neither Kuhn nor Laudan claimed to derive specific norms. Stanford does. Yet absent causal evidence, such haste to effect policy is often unwarranted; after all, the attribution of causality is "a *sine qua non* when applying empirical analysis to public policy" (Tirole, 2017, 118). Thus, it is actually a virtue of Kuhn's and Laudan's arguments that they opt for vague hints over specific conclusions.

Indeed, it should have been obvious that Stanford's move to policy was unwarranted. When drafting policy, we should not only be able to measure the variables of interest; more pressing still, we should know *when* we can measure them. Yet when do we test for realism or instrumentalism, or for their supposed effects? Stanford's case studies, and even his more recent work, are silent on this. Yet as Levy (2015, 398) notes, "our confidence in the validity of counterfactual predictions is a function of the temporal distance and length of the causal chain from antecedent to consequent." Essentially, the more distant a temporal proximity, the more mechanisms can interact; consequently, the greater the uncertainty of the counterfactual. Hence, the more you must consider factors exogenous and endogenous to the counterfactual that might redirect you to the actual outcome (so-called redirecting counterfactuals). Stanford did nothing of this sort. Earlier, in §2.1, I skipped over these criteria (*temporal proximity* and *redirecting counterfactuals*) because Stanford's argument violates them almost by definition: by using alternatives that were *later* accepted—indeed, often *much* later—to argue for the existence of UAs, short (hence manageable) time scales are almost invariably excluded.

Being trapped in this dilemma tells us that science's history, by Stanford's own lights, has failed to deliver his goal: the scientific realism debate still does not concern a difference that makes a difference.

2.4.2 Hasty abstraction is the problem, not history

All the same, I do not think Stanford failed because of anything peculiar to historical evidence; rather, I argue he failed because philosophers of science too hastily abstract from concrete concerns. In short, philosophers of science too often begin by abstracting without a clear goal, then attempt to derive predictions from that abstraction; this is in contrast to identifying a clear goal, then making and testing predictions, and abstracting as necessary to achieve that goal. Stanford unapologetically begins with a series of abstractions, seemingly because they are traditional in philosophy of science (Stanford, 2020, 17–8):

> Since its inception in the writings of Smart, Putnam, Van Fraassen, Boyd, Laudan and others, the modern realism debate has been predicated on the assumption that there is some *point* to ascending to the levels of abstraction at which we generalize about our "mature scientific theories" and their "empirical successes" or "approximate truth." For realists, the point of that ascent was to try to explain the success of such theories in a way that revealed broad epistemic categories that would allow us to reliably pick out collections of scientific theories or other privileged parts of contemporary theoretical orthodoxy that we can justifiably treat as secure epistemic possessions and can therefore be trusted to persist in some rec-

> ognizable form throughout the remaining course of scientific inquiry itself. The instrumentalist thinks we learn quite different lessons from considering matters at this level of abstraction and generality: she is convinced by the historical evidence, for example, that we should expect many of even the most fundamental commitments of contemporary theoretical orthodoxy to be eventually overturned and that there are no such general epistemic features or categories we might use to form reliable expectations regarding which of those commitments will or will not be preserved in some recognizable form throughout the course of our further scientific investigation of the world. Of course, when the instrumentalist claims that we should believe only a successful theory's empirical implications, or what it says about observable matters of fact, or some such, she too is offering a (competing) abstract and general criterion of epistemic security for our scientific beliefs.

Here Stanford asserts that there is a point to ascending to these levels of abstraction, and it is implied that this point speaks in favor of ascension. One might quibble over whether this point really speaks in favor of ascension. This is a red herring, however. In fact, the "point" of ascension here is one in name only, for it is entirely unconstrained by any practical, concrete goal. Baldly put, the abstractions of the scientific realism debate are pointless.

Contrast the realism debate abstraction with one I claim is constrained by a concrete goal: Kate Manne's definition and account of misogyny. In Stanford's phraseology, Manne ascends to the levels of abstraction at which we generalize about our "social environments" and their "policing" and "enforcing" of gendered norms and expectations, to the detriment of girls and women Manne (2020). Manne describes the point of her ascension as follows Manne (2020):

> On the whole, though, my account of misogyny counsels us to focus less on the individual *perpetrators* of misogyny, and more on misogyny's *targets* and *victims*.

> This is helpful for at least two reasons. First, some instances of misogyny lack any individual perpetrators whatsoever; misogyny may be a purely structural phenomenon, perpetuated by social institutions, policies, and broader cultural mores. Second, understanding misogyny as more about the hostility girls and women *face*, as opposed to the hostility men *feel* deep down in their hearts, helps us avoid a problem of psychological inscrutability. It's often difficult to know what someone's innermost states and ultimate motivations are, short of being their therapist (and even then, such knowledge may be elusive). But my account of misogyny doesn't require us to know what someone is feeling, deep inside, in order to say that they are perpetuating or enabling misogyny. What we need to know is something we are often in a much better position to establish: that a girl or woman is facing disproportionately or distinctively gendered hostile treatment because she is a *woman in a man's world*—that is, a woman in a historically patriarchal society (which includes, I believe, most if not all of them). We don't need to show that she is subject to such treatment because she is a woman *in a man's mind*—which, in some instances, can't be the issue.

Obviously, this abstracts away from the beliefs and behaviors of individuals, which are the concern of the typical definition. Nevertheless, this abstraction has a serious, concrete "point" insofar as it serves a concrete goal. We can see this in two ways. First, the abstraction does not shift the goal. That is, all entrants to the debate still share an understanding of the overall goal, feminist or not: to understand and address the hostile treatment girls and women (disproportionately) face simply because of their gender. Indeed, presumably feminists and non-feminists both understand especially egregious acts, like Elliott Rodger's femicidal spree, as hostile treatment toward women that should be eliminated. This is true even if the non-feminists do not understand other aspects of society to be hostile toward women and girls. Second, the point of ascension is clear: because the goal is to understand

and address the hostile treatment girls and women *face*, not merely the hostile treatment men (knowingly) *perpetrate*, abstracting away from individuals captures more of the cases we care about. Indeed, you can easily put this abstraction to use to draw together a variety of phenomena, as Manne does, without significantly obscuring the relevant details of individual forms of misogyny—all while furthering our ability to address actual hostile treatment of girls and women. In short, Manne's is an abstraction that clearly supports a concrete goal.

The realism debate abstraction is not constrained by a concrete goal in this way. Indeed, it is even difficult to see how the abstraction's "point" could serve a concrete goal. First, the two sides of the realism debate do not predicate ascension on the *same* point. Absent this, it is difficult to see how they can have the same concrete goal, for typically opponents sharing a goal *and* a method also see the same point for using the method. Of course, the realist and instrumentalist do both ascend in search of an "abstract and general criterion of epistemic security." To be sure, this is a *shared* goal. It is clearly not a goal *preceding* ascension, however, hence it is not a goal with respect to which ascension has been deemed an appropriate method. For one, few who develop and use individual theories discuss them using abstruse terms like "approximate truth" or "empirical success," absent invocation of the abstract debate itself; in fact, these terms seem to have been invented *after* ascension as a way to corral all of the actually used (and hence useful) terms. For another, the realist and instrumentalist do not precede the ascension by jointly specifying concrete theories that stand to benefit from this ascension. Indeed, before ascending, neither identify *any* particular theory whose epistemic securing is a goal. At any rate, if there were a concrete goal, they would be searching for means to that shared end, not an end for their shared means. Thus, the abstraction itself has altered the goal.[22] Second, foregoing a shared goal, it is not even clear what concrete goal the realist's (resp. instrumentalist's) individual "point" serves. Why for instance, should ascension be beneficial to the realist? What is their scientific goal,

[22]Of course, changing one's goal after an abstraction is not inherently bad. Nonetheless, one should be able to say how the goal has changed and why. That has not occurred here.

or what concrete problem do scientists face, such that ascension is appropriate? We can speculate, of course. Yet were the "point" of ascension tied to a concrete goal, we shouldn't need to speculate: the point would naturally serve the goal, as it did for Manne. Hence, the realism debate abstraction is pointless; the debate has no goal by which to determine whether it has a point.

Curiously, the instrumentalist's lesson from the abstraction pushes us back toward the concrete, too. The instrumentalist's "point" in abstracting, roughly, was to show that we cannot (and should not try to) predict what features of our current scientific theories will persist. Suppose this is straightforwardly correct. Stanford claimed this is supported by the history of science, particularly by the case of Darwin.[23] Yet, one of the specific lessons Stanford seemingly took from Darwin's case is that unchallenged upstream inferences are often to blame for downstream troubles. And certainly part of what revealed those upstream errors was a more rigorous collection of data and a more prudent evaluation of cause and effect. But surely this applies also to theories *of* science! If this is right, then it should be no comfort to us that the likes of Smart, Putnam, Van Fraassen, Boyd, Laudan and others assumed abstractions like "mature scientific theories", "approximate truth", etc., would persist—for what other terms could we possibly use?—despite subsequent changes downstream. Comparing to Darwin, we can reasonably expect at least part of the treatment for this ill-advised upstream inference to involve remaining at more concrete levels of abstraction and rigorously collecting and analyzing data that concern us there. Then, *if necessary*, we might again ascend to higher levels of abstraction.

Naturally, what it means for would-be realists and instrumentalists to "remain at more concrete levels of abstraction" will vary according to what they care about most, i.e., what their goal is. A few examples may help. For one, Stanford himself appears to care most about peer review. In this case, "remaining" could mean participating in reform movements

[23]Of course, precisely what is being supported and how is unclear, as we glean from the troubles encountered in §2.

that scientists of science have already undertaken, like the push for open peer review (see, e.g., (Ross-Hellauer, 2017) and (Tennant et al., 2017)). Others appear to care more about providing (self-consciously defeasible) predictions about the shape of our near-future theories, in the hopes that pursuing these shapes will bring new insights. Thus "remaining" could involve identifying and disambiguating (e.g., (Schneider, 2019)) or even generating (e.g., (Williams, 2019)) predictions about future theories. Perhaps, too, a would-be realist or instrumentalist cares about the psychological and sociological aspects of knowledge and understanding. "Remaining," then, could also mean taking part in areas of inquiry like the psychology of science (see, e.g., (Proctor and Capaldi, 2012)), science as psychology (see, e.g., (Osbeck et al., 2011)), or social epistemology.

"Remaining concrete" in any of these guises does not mean giving up on philosophy of science, and, indeed, it seems likely that philosophical and historical training would be a boon. However, it does appear to mean giving up on the abstractions of the realism debate, at least for now, since arguing at the level of "abstract and general criteria of epistemic security" promises to erase complexities important for more concrete goals. Yet this should not be surprising to anyone familiar with establishing or evaluating norms. As Cartwright (2012) shows, there is often a tradeoff between generality and relevance when it comes to making policy on the basis of evidence from randomized-control trials. Why should this lesson not hold when the evidence comes from science's history? History, in this respect, is not responsible for the failure of HIN philosophy of science. Rather, philosophy of science is to blame for the failure of the HIN method. Thus, if the point of philosophy of science is to articulate and defend norms for science, regardless of whether it uses history via the HIN method, it must be more concrete.

2.5 Conclusion

In this paper I have argued that using science's history as evidence has not improved the quality of normative philosophy of science. I have shown that Stanford's argument for adopting *Instrumentalism*—via a historical argument for *Historicism*—is fundamentally flawed. This placed Stanford in a dilemma: either he is no different from his HIN predecessors, or he has no evidence for his preferred norm. I traced this dilemma to the scientific realism debate's presumption that there is a point to talking abstractly about "mature scientific theories", "approximate truth", etc. I then argued that this abstraction is pointless because the debate has no goal. Finally, I showed that not having a concrete goal for abstraction is responsible for the failure to derive useful norms from the scientific realism debate—i.e., that the debate is the problem, not history. This means that the scope of my criticism is not just HIN philosophy of science but normative philosophy of science more generally. Thus, I conclude that if the point of philosophy of science is to develop norms that benefit practicing scientists and society, it must be more concrete.

Chapter 3

The (Not So) Hidden Contextuality of von Neumann's "No Hidden Variables" Proof

3.1 Introduction

Von Neumann's (in)famous proof of the non-existence of hidden variables in quantum mechanics is commonly discussed in the context of work that came much later, namely that of Bohm and Bell. In this context, the goal of specifying a set of axioms is to identify *only* what is essential for *any* quantum theory. Call this axiomatic reconsideration. Thus, given that Bell (rightly) criticized one of von Neumann's assumptions, the story goes that von Neumann made a grave error; even worse, von Neumann thereby *erroneously* claimed to have ruled out hidden variables.

This story is wrong—or, so I argue. However, in the main, I do not disagree either on the historical or the physical facts. Indeed, excellent exegetical work has already been done on von Neumann's work in physics (Duncan and Janssen, 2013) (Lacki, 2000) (Rédei, 1996) (Rédei, 2006) (Rédei and Stöltzner, 2006) (Stöltzner, 2001), including on his no hidden variables proof (Bub, 2011) (Bub, 2010) (Dieks, 2017) (Mermin and Schack, 2018) (Stöltzner, 1999). Instead, my disagreement concerns the framing, which lumps von Neumann in with Bohm and Bell (especially the latter). Here I argue that von Neumann was performing an *axiomatic completion* of quantum mechanics, where 'quantum mechanics' refers to a *specific* theory of quantum phenomena rather than, vaguely, to *any* theory of quantum phenomena.[1] This axiomatic completion relied on Hilbert's axiomatic method. With this understanding at hand, I re-interpret the history of von Neumann's no hidden variables proof.

The argument proceeds as follows. In the first section, I give an overview of the axiomatic endeavors foreshadowed in (Hilbert et al., 1928). Here I emphasize three features of the Hilbertian axiomatic method: (1) it requires the separation of the *facts* of a given theory from the *formalism*, (2) the formalism is *uniquely* characterized with respect to the theory, and (3) the goal was to *order* the area of knowledge and *orient* its further research. As such, the method's results were *provisional* and *relative*. In the second section, I describe the history of quantum theory that immediately preceded (Hilbert, Neumann, and Nordheim 1928). Here I focus on the influence and status of the transformation theory as developed by Dirac and Jordan. In the third section, I re-interpret von Neumann's work in 1927 as the axiomatic completion of quantum mechanics, where the latter is understood as the work coming out of the Göttingen—Cambridge tradition. Here my central claim is that insofar as von Neumann very likely already had his "no hidden variables" proof in 1927, he had thus demonstrated that the Hilbert space formalism was the *unique* representation of quantum mechanics. I also introduce a little-known debate between von Neumann and Schrödinger on the status of hidden variables. In the fourth section, I show that von Neumann's 1932 book made his use of the axiomatic method—including its character as provisional and relative—

[1] In what follows I will use 'quantum theory' to refer to what we today call 'quantum mechanics' and reserve 'quantum mechanics' for its historical referent, i.e., the cluster of work that grew up in Göttingen and Cambridge.

explicit; in this sense, *nothing was deeply hidden* concerning his motivations. In the fifth section, I briefly revisit the infamous proof of IV.1 and IV.2 to show how it is to be read as an axiomatic completion. Finally, I conclude by discussing the legacy of von Neumann's axiomatic completion of quantum mechanics insofar as it oriented later inquiry.

3.2 The Axiomatic Method: Separating Facts from Formalism

A common misconception of Hilbertian axiomatics holds that it promoted formalization of scientific and mathematical theories in the service of radical epistemological or metaphysical goals. Wilson (2017, 151), for instance, has repeatedly suggested that the program intended to reveal the basic metaphysics of theories, analogous to later intentions for rational reconstruction. Similarly, Lacki (2000, 315) characterizes Hilbert's interest in axiomatization as residing "in his care for logical clarification and rational reconstruction." Yet this view trades on half-truths. Closer inspection reveals a richer and more grounded program in which Hilbert is not so naive, von Neumann not so facetious, and the labor not so fruitless.

The core feature of the axiomatic method comes in Hilbert's maxim to "always keep separated the mathematical apparatus from the physical content of the theory" (Lacki, 2000, 313).[2] Concerning quantum theory in particular, we find an expression of this in the following

[2]Lacki seemingly understands this to be synonymous with the imperative to delimit "as close as possible[...]what are the minimal assumptions on which to secure its [quantum theory's] foundations, assumptions which should be sufficiently beyond any doubt so that one could consider them safely as not subject to further revision" (Lacki, 2000, 313). In a follow-up to this paper, I argue that this conflates two axiomatic spirits, those of axiomatic reconsideration and axiomatic completion, and that the former is an anachronism during this period. Note that were the axiomatic method to demand "assumptions sufficiently beyond any doubt," then Hilbert's axiomatization of geometry would have been an abject failure. Here, it will become clear that we need not assume von Neumann was any different from Hilbert on this.

passage from (Hilbert et al., 1928), which remarks on the ideal (non-obtaining) way to a quantum theory[3] (italics mine; page references are to (Von Neumann, 1963, 105)):

> The way to this theory is as follows: Certain physical demands on these probabilities are suggested by our past experiences and trends, and their satisfaction necessitates certain relations between the probabilities. Secondly, one seeks a simple analytic apparatus in which quantities occur that satisfy precisely the same relations. This analytic apparatus, and with it the operands occurring in it, now undergoes a physical interpretation on the basis of the physical demands. The aim in doing so is to so fully formulate the physical demands that the analytic apparatus is uniquely defined. This way is thus that of an axiomatization, as has been carried out, for example, with geometry. Through the axioms the relations between the elements of geometry, point, line, plane, are characterized and then it is shown that these relations are exactly satisfied by an analytic apparatus, namely the linear equations.
>
> In the new quantum mechanics, one formally assigns a mathematical element, which is in the first instance a mere operand, as representatives according to a certain specification of each of the mechanical quantities, but from which one can receive statements about the representatives of other quantities and thus, through back-translating, statements about real physical things.
>
> Such representatives are respectively the matrices in the Heisenberg, the q-numbers in the Dirac, and the operators in the Schrödinger theories and their present developments.
>
> *It is therefore important to note that we examine two wholly different classes of things, namely on one hand the measurable numerical values of physical quanti-*

[3] I quote at length because the passage is not readily available in English. All translations are mine, unless specified.

> *ties and on the other their assigned operators, which are calculated with strictly according to the rules of quantum mechanics.*
>
> The above suggested procedure of axiomatization is not typically followed in physics now, but rather is the way to the erection of a new theory, as here, according to the following principles.
>
> More often than not, one supposes an analytic apparatus before one has yet specified a complete system of axioms, and then arrives at the establishment of the basic physical relations only by interpretation of the formalism. It is difficult to comprehend such a theory when one cannot sharply distinguish between these two things, the formalism and its physical interpretation. This divorce should here be made as clearly as possible, when we also, in accordance with the present state of the theory, don't yet wish to found a complete axiomatics. In any case, what is certainly well-situated is the analytic apparatus, which—as purely mathematical—is also capable of no modification. What can, and probably will, be modified about it is the physical interpretation, with which exists a certain freedom and arbitrariness.

The bigger picture here is rather clear: while quantum mechanics took the non-ideal path to its current position, during which the physical facts[4] and formalism were not clearly separated, an ideal erection of the theory would cleanly separate the two. Indeed, though the axiomatic method and this separation maxim are today associated with Hilbert, they were more routinely recognized in the day. For instance, this is the feature of the axiomatic method that Einstein's "Geometrie und Erfahrung" identifies as the solution to the "riddle" of how

[4] I refer to 'physical facts' rather than 'physical interpretation' throughout to avoid conflation with our modern notion of philosophical interpretation, which relies on a pseudo-model-theoretic understanding of theory—world relations that was likely unavailable at the time Eder and Schiemer (2018) and, at any rate, is not compatible with Hilbert's above description of his axiomatization of geometry. Besides, 'physical facts' better conforms to Hilbert's account (especially "Axiomatische Denken" (Hilbert, 1917)) as well as the broader physics community's account (e.g., "Geometrie und Erfahrung" (Einstein, 1921)) of axiomatization.

mathematics—whose objects are purely imaginary—can apply to actual objects (Einstein, 1921, 3–4):

> So far as the propositions of mathematics correspond to reality, they are not certain, and so far as they are certain, they do not correspond to reality. Complete clarity on the situation seems to me to have come into the community's possession only through the method of mathematics known by the name of "Axiomatics." The progress achieved by the axiomatic method consists in the fact that it cleanly separates the logical-formal from the factual or intuitive content; only the logical-formal is the subject of mathematics according to the axiomatic method, but not the intuitive or other content connected to the logical-formal.

Thus, in broad strokes, the axiomatic method is for separating the mathematical (logical-formal) from the non-mathematical (intuitive) content.

This tells us not only *what* the axiomatic method is for but also *whom* it is for: the meta-mathematician. We can spell this out using the example of Hilbert's axiomatization of geometry, to which Hilbert referred above.[5] In brief, a more-or-less agreed upon set of propositions was taken to constitute Euclidean geometry at the end of the 19th Century. However, it was unclear just what assumptions were necessary to derive these propositions. In particular, it was asked whether Archimedes' Axiom—i.e., for any previously given line segment CD, every line segment AB can be repeated such that the length of that segment exceeds CD—was necessary:[6] while it was clear that the proposition was true on our "in-

[5]The account of Hilbert's axiomatic method that follows is my own. However, it is similar in important respects to especially (Baldwin, 2018, chap. 9), (Detlefsen, 2014), (Peckhaus, 2003), and (Corry, 2004), as well as (Hallett, 1990)(Hallett, 1994)(Hallett, 2008), (Sieg, 2014), and (Wilson, 2020). For entry into interpreting Hilbert's *Grundlagen der Geometrie*, see (Giovannini, 2016) and (Eder and Schiemer, 2018). For an account that emphasizes more of the foundationalist aspects of the axiomatic method, based on Hilbert's work related to general relativity, see (Brading and Ryckman, 2008)(Brading and Ryckman, 2018), as well as (Brading, 2014) and the editors' remarks in (Sauer and Majer, 2009).

[6]I discuss Archimedes' Axiom here to avoid some of the messiness of the history of the parallel postulate. However, I take it that the same point applies to the parallel postulate—e.g., (Eder and Schiemer, 2018, 66–7)

tuitive" conception of Euclidean geometry, its formal relation to the other axioms had not yet been clarified. Enter Hilbert, the meta-mathematician. Taking the more-or-less agreed upon propositions (constituting Euclidean geometry) as the targets for recovery—but eschewing the "intuitive" moorings Euclidean geometry had accumulated in its practical go of it!—Hilbert characterized the axioms necessary for recovering these target propositions. Crudely put, where the mathematician asks what propositions *a given set of axioms* suffice to prove, the meta-mathematician turns this around via the axiomatic method to ask what axioms are necessary to prove a given set of propositions. Thus, the axiomatic method is primarily for the meta-mathematician.

Before considering quantum mechanics, I want to highlight a couple features of the axiomatic method using Hilbert's Grundlagen der Geometrie as an example. These features are easiest to draw out from Hilbert's "Axiomatische Denken" (Hilbert, 1917). Firstly, the method is *relational* in two senses. On the one hand, the method is relational in that the formal theory is given relative to a more-or-less comprehensive field of knowledge, which is itself constituted by facts. The facts of such a field of knowledge admit of an ordering that (Hilbert, 1917, 405)

(italics mine): "...lacking the precision of an exact axiomatization and a methodologically clean understanding of what is at stake when we ask ourselves about the independence of the axiom of parallels, these results [i.e., non-Euclidean geometries] were still hotly debated among philosophers. This is certainly due in part to the empirical content people associated with geometry and the fact that matters of logical consequence were mixed up with matters of empirical truth." Analogous to the body text, then, my point is that axiomatizing the full set of empirical truths (as then understood) allowed for the eventual trimming of the axiom of parallels precisely because the axiomatization laid bare its logical consequences in the context of the other axioms.

> is effected in each case with the help of a certain truss [Fachwerk][7] of concepts in such a manner that to each individual subject of a field of knowledge corresponds a concept from this truss and to each fact within the field a logical relation between the concepts. The truss of concepts is none other than the theory of the field of knowledge.

Thus the conceptual truss plays a crucial role in the axiomatic method. Nevertheless, Hilbert is at pains to stress that the theory is *not* the same as the field of knowledge; rather, we quickly gather, as here concerning consistency and independence, that the theory plays a precise, practical role insofar as it *represents* that field of knowledge (Hilbert, 1917, 407)(bold added):

> Should the theory of a field of knowledge—i.e., the truss of concepts whose goal is to represent it—serve its express purpose of **orienting** and **ordering**, then it surely must meet two standards especially: *firstly* it should provide a survey of the dependence resp. independence of the propositions of the theory and secondly a guarantee of the lack of contradictions among the propositions of the theory. In particular the axioms for each theory are to be examined from these two perspectives.

To put it crudely and anachronistically, Hilbert understands theories as "lossy" representations of fields of knowledge. In putting it this way, I mean to stress that Hilbert was not

[7] While typically translated as "framework," I think its meaning is better captured by "truss." In English-language philosophy these days, "framework" and "system" are often taken to be synonymous, hence we typically assume that frameworks are fairly fleshed-out or robust affairs. But I think this is a mistaken assumption in German, and particularly here: if we instead understand "Fachwerk" as more like "truss"—as a first-pass support, upon which a more robust framework is built—then we may fairly assume that a "Fachwerkes der Begriffe" is more of a stepping-stone on the way to a system of axioms, meaning that concepts (in their informal state) go through a vetting process before being precisified in an axiomatic system. My intention in drawing this distinction is to highlight that the concepts are not necessarily the invention of the (mathematical) theoretician, but rather are often provided by the scientists or other experts of the area in question, and they are merely codified more rigorously by the mathematician. (Obviously, rigor in the spirit of Hilbertian axiomatics.)

so naïve to think that the "ladder" of facts is "kicked away" once a theory has been given. Rather, he recognized that one has inevitably made non-trivial choices in how to represent these facts, regardless of how well-informed such choices are.

Concerning Hilbert's *Grundlagen*, the picture is something like the following. First, the facts are the "more-or-less agreed upon propositions" that were taken to constitute Euclidean geometry. Insofar as it is in this set of propositions, this includes Archimedes' Axiom. These facts admit of an ordering, given in Hilbert's *Grundlagen* by the choice of axioms. This ordering is effected "with the help of [the] truss of concepts" that inform the groupings Hilbert gives to the axioms (namely, connection, order, parallels, congruence, and continuity). By aligning the subjects of Euclidean geometry with concepts and its with logical relations among those concepts, this truss of concepts constitutes a *theory* of Euclidean geometry. Hilbert then examines the independence and consistency of the various groups of axioms.

This brings us to the way in which, on the other hand, the method is *relational* in the more traditional mathematical sense: independence and consistency results are relative to some prior mathematical theory. This is essentially a response to a problem I skirted a moment ago when introducing Hilbert's *Grundlagen*. There I noted the importance of eschewing the "intuitive" moorings a set of facts has accumulated in its practical go of it. This is far easier said than done, however. In the *Grundlagen*, as Hilbert et al. said above, Hilbert used the linear equations as an analytic apparatus that represented precisely the relations arising in his theory of geometry. By translating unresolved geometric questions into the better-understood language of linear equations, Hilbert was able to answer these questions in the latter and then translate their answers back into geometry. Doing this he, for instance, reduced the question of the consistency of the geometric axioms to the question of the consistency of the theory of analysis.

That said, there is an extra trick here, which is that the analytic apparatus needs to be uniquely specified with respect to the physical facts as codified by the truss of concepts.

This is only implicit in "Axiomatische Denken," but it is explicit in (Hilbert et al., 1928). There they say that the axiomatic (ideal) way to a (quantum) theory is to specify the physical demands on probabilities, then develop an analytic apparatus "in which quantities occur that satisfy precisely the same relations," and finally "interpret" the apparatus on the basis of the physical demands. One might hastily conclude that the relationship between apparatus and interpretation is the one familiar to us, namely one where the hope is to have so specified the mathematics that all interpretations are equivalent in some model-theoretic sense. Precisely the opposite is desired here, however: "The aim in doing so [interpreting the analytic apparatus] is to so fully formulate the physical demands that the *analytic apparatus* is uniquely defined" (italics added). And this should make sense. If a set of physical demands admits of multiple formalizations, then claims to have separated the factual from the formal in the area are dubious. In other words, Hilbert's axiomatization of geometry needed a 1-1, invertible mapping between the languages of geometry and analysis in order to effect the translational strategy, i.e, replacing questions of geometrical truth with arithmetical truth. With this in hand, Hilbert can confidently claim to have represented the truths of geometry as truths of arithmetic. So, the axiomatic method applied to the area being investigated demands a, presumably better understood, analytic apparatus to which the area's theory corresponds uniquely.

The upshot is that the axiomatization of an area is *relative* both to that area's set of facts and to an analytic apparatus, and the latter relation demands a unique correspondence between the theory of the area and the analytic apparatus. This brings us to the second feature of the axiomatic method. This is that it is a continual process whose results are always *provisional*, especially insofar as they are non-mathematical. Yes, fundamental propositions of the theory can be viewed as axioms of the area of knowledge (Hilbert, 1917, 406):

> The fundamental propositions can be seen from an initial standpoint as the axioms of the individual field of knowledge: the continuing development of the

> individual field of knowledge is then founded merely on the further logical development of the already-presented framework of concepts. This standpoint is prevailing especially in pure mathematics, and we owe the corresponding method of working for the tremendous developments of geometry, arithmetic, function theory, and the whole of analysis.

Temptation would have us conclude that axioms are not subject to revision, and above Hilbert et al. did say that the *analytic apparatus* of quantum mechanics was "well-situated" and "capable of no modification." But this is not a very interesting claim: indeed, formalisms *are* certain in that they never *require* modification—but only so far as they are mathematical, as Einstein said above. And while we hear this, too, in "Axiomatische Denken", he at the same time stresses that axioms are always provisional owing to their function (Hilbert, 1917, 407) (italics mine):

> Consequently, in the mentioned cases, the problem of the grounds of an individual field of knowledge had then found a solution; however, this was only provisional. In fact the need asserted itself in the individual sciences mentioned to in turn ground the propositions looked at as axioms which lay at the foundation. Thus one reached "Proofs" of the linearity of the equation for the plane[...], of the law of entropy and the proposition of the existence of roots of an equation.
>
> But the critical examination of these "Proofs" affords recognition that they are not themselves proofs, but rather at those depths merely make possible the tracing back to certain deeper-lying propositions, which henceforth are to be seen for their part as axioms in place of the propositions to be proven. Thus originate the axioms actually so-called today in geometry, arithmetic, statistics, mechanics, radiation theory or thermodynamics. These axioms form a deeper-lying layer of axioms with regard to the one layer, as they have been characterized through the propositions just mentioned as lying at the ground in individual fields of

> knowledge. The procedure of the axiomatic method, as it is characterized here, hence comes to a deepening of the foundations of the same individual field of knowledge, of course to the extent such is necessary with each framework, when one elaborates the same field wishing to go higher and yet vouch for his security.

But recall that the primary benefit of axiomatization is separation of factual and formal content. Thus, if the axiomatic method is indefinitely applicable—in the sense that axioms so-called today could be replaced tomorrow—it is not hard to see that "going deeper" could rule today's so-called formal content as factual tomorrow, or perhaps even vice versa. Indeed, if this were not possible, it is difficult to see how a deepening of the foundations—to which the procedure of the axiomatic method comes—could vouch for the security of one who wishes to elaborate on a formalism: there would be no "insecurities" (=hidden factual content) to remove. We gather, thus, that Hilbert does not consider axioms as beyond revision, *even in pure mathematics*.[8] Rather, there is an expectation that deeper layers will be found when the need for them arises, and this inevitably will lead to the demotion of axioms today-so-called in favor of deeper-lying ones.

All told, then, axiomatization is a *relational* and *provisional* method for separating between the factual and formal content of an area of knowledge. It is relational first in the sense that it is relative to a chosen set of facts from that area of knowledge and, at the same time, a truss of concepts from which one constructs a theory of that area. Second, it is relational in the sense that independence and consistency results for the theory are relative to an analytic apparatus that is uniquely picked out by the strictures of that area's theory. The method is provisional in the sense that the axioms it produces to represent of an area of knowledge are subject to revision as our understanding of that area changes.[9] Finally, an axiomatization

[8] However, this does not mean that revising a theory involves denouncing the previous one. In such a case, we have simply shifted our interests, and the previous theory is still perfectly acceptable as a mathematical theory. (This line on Hilbert gets trickier to defend when it comes to elementary arithmetic or logic, but this needn't concern us here.)

[9] Likewise for consistency and independence results, since we may also revise what we consider the canonical translation. Historically, this just seems less likely to change. That said, I suspect this is the

aims to *orient* and *order* our inquiry in an area of knowledge, especially our mathematical inquiry, by focusing our attention on what is central to its theory.

3.3 The (Pre-)History of Hilbert, von Neumann, Nordheim (1928)

As (Hilbert et al., 1928) was being written—in early 1927—quantum theory faced several problems. In the previous two years, there had arisen not one, but two[10] calculational techniques for predicting quantum phenomena: matrix (quantum) and wave (undulatory) mechanics.[11] Each had met with some predictive success. However, the two calculational techniques appeared fundamentally different on their face. In fact, the two theories were then known to differ in rather significant ways, both mathematically and physically. To put it mildly, quantum theory was a mess. In arguing that the two were not, in fact, equivalent, Muller (1997a, 38) (Muller, 1997b) (Muller, 1999) summarizes the distinctions as follows:

> One reason for the failure of the mathematical equivalence is the fact that whereas matrix mechanics could in principle describe the evolution of physical systems over time (by means of the Born-Jordan equation), but limited itself unnecessarily to periodic phenomena, wave mechanics could not—Schrödinger's time-dependent wave-equation dates from 3 months later than his equivalence proof. Other reasons for the failure of mathematical equivalence are: the absence in ma-

better way to understand Hilbert's "proof" of the Continuum Hypothesis against the backdrop of the later work of Gödel and Cohen: Hilbert, thinking syntactically, assumed his L-like structure exhausted the truths of set theory; Gödel and Cohen, thinking semantically, showed that this assumption was non-trivial.

[10]Really, four, but I will follow the usual convention of ignoring Born and Wiener's operator mechanics and Dirac's q-numbers.

[11]The usual list of Göttingen matrix mechanics publications includes (Heisenberg, 1925) (Heisenberg, 1926) (Born et al., 1925) (Born et al., 1926); English translations for three of these can be found in (van der Waerden, 1967). The usual list for wave mechanics includes Schrödinger's four "Quantisierung als Eigenwertproblem" papers and (Schrödinger, 1926), which can all be found in (Schrödinger, 1927a) (English translation: (Schrödinger, 1927b)).

> trix mechanics of a state space but its presence in wave mechanics (the space of wave-functions); the fact that Euclidean space and a set of charge-matter densities, both prominently present in wave mechanics, had no matrix-mechanical counterparts; and the fact that matrix mechanics produced the first theory of a quantised electromagnetic field by means of matrix-valued fields, whereas Schrödinger emphasised there was no need to tinker with the classical Maxwell equations in wave mechanics.

None of these differences were hidden from view, and their most significant difference—the apparent discreteness of matrix mechanics versus the apparent continuity of wave mechanics—was often discussed.

Yet despite these differences, the two calculational techniques led to the same answers in a number of elementary problems. This was considered a promising development by many. Schrödinger himself, in addition to Sommerfeld, was quickly convinced that wave mechanics was equivalent (or at least, made matrix mechanics superfluous) (Mehra and Rechenberg, 1987, 638–9). As he wrote in a 22 February, 1926, letter to Wien (translation by Mehra and Rechenberg):

> Relation to Heisenberg. I am convinced, along with Geheimrat Sommerfeld, that an intimate relation exists. It must, however, lie rather deeply, because Weil [sic], who has studied Heisenberg's theory very thoroughly and has developed it further himself, says upon reading the first manuscript that he is unable to find the connection. Consequently, I have given up looking any further myself....Now I firmly hope, of course, that the matrix method, after its valuable results have been absorbed by the eigenvalue theory, will disappear again.

(Note, too, that Schrödinger here refers to his theory as an eigenvalue theory. This will be important momentarily.) Despite telling Wien that he had given up looking for the connection, we know he did not. Not long after—somewhere between one and four weeks later—Schrödinger managed to prove that, in a certain respect, wave mechanics was equivalent to matrix mechanics.[12]

Two things are important here. First, it bears repeating that Schrödinger was not only *convinced* almost immediately that his wave mechanics could absorb matrix mechanics, but he strongly *desired* its absorption. Second, he made good on this claim (he thought) precisely by *expanding* wave mechanics to *make* it equivalent (Muller, 1997a, 54–5):

> In his second founding paper Schrödinger had confessed that wave mechanics could not calculate the intensity of spectral lines, something which matrix mechanics in principle was able to do. But from his own proof [of equivalence] Schrödinger learned how to express spectral-line intensities in wave mechanics, by looking at the [relevant] matrix-mechanical formula (6). ...So besides extending wave mechanics by adding the canonical wave operators (Postulate W1) whilst in the process of proving equivalence, Schrödinger was here extending wave mechanics once more by another brand new postulate. Schrödinger was not just attempting to prove the equivalence of matrix mechanics and extant wave mechanics, but he was also *expanding* wave mechanics on the spot to *make* it equivalent to matrix mechanics.

These are important to stress because they complicate any claim that Schrödinger was coerced or otherwise unduly influenced into accepting the quantum hegemony. Unknowingly or not, his *choice* to make wave mechanics equivalent to matrix mechanics—even in so limited

[12]See (Perovic, 2008) on the goal of Schrödinger's proof. As there noted (Perovic, 2008, 459), von Neumann takes Schrödinger to have demonstrated the mathematical equivalence of the two theories, contrary to what Perovic claims Schrödinger was after.

a setting as Bohr's model of the atom—set him on the path to hegemony.[13] In what follows, I will stress Schrödinger's own choices and views to this end.

In the wake of matrix and wave mechanics there arose the transformation theory, developed by Jordan, Dirac, and London. The transformation theory brought with it three things that are significant here. First, it resolved lingering questions concerning the relationship of the various quantum calculi: they were all equivalent, as far as the transformation theory was concerned. Mehra and Rechenberg, in fact, conclude their discussion of the transformation theory by quoting what Oskar Klein later told Kuhn: the transformation theories of Jordan and Dirac "were regarded as the end of the fight between matrix and wave mechanics, because they covered the whole thing and showed that they were just different points of view" (Mehra and Rechenberg, 2001, 89). Unsurprisingly, the language physicists used changed, too, so that 'quantum mechanics' came to refer not just to matrix mechanics but also to those calculi captured in the transformation theories, as well as the transformation theories themselves. This is seen already in the early presentations of transformation theory by (Jordan, 1927, 810) and (Dirac, 1927, 621), each of whom: uses 'quantum mechanics' to refer loosely to the various calculi (but clearly not intending to capture any unconceived alternative "theories" of quantum phenomena); refer to Heisenberg's quantum mechanics instead as 'matrix mechanics'; and call wave mechanics, for instance, a "representation" (resp. for Jordan, "Form"). Thus, the theory of *quantum mechanics* is the one arising specifically from transformation theory.

Second, the transformation theory replaced the morass of interpretation-adjacent mathematical questions plaguing the various forms of quantum mechanics with essentially one. Where before matrix and wave mechanics faced related but distinguishable questions about the validity of their calculi's methods, transformation theory faced instead the single question

[13] For one, as Muller notes, Schrödinger is now committed to the energy basis as physically preferred because the postulate he provides—roughly, that the spectral-line intensities in wave mechanics are proportional to those of matrix mechanics—fails for every other basis.

of the domain of validity of Dirac's delta function. This is reflected in (Hilbert et al., 1928) (Von Neumann, 1963, 105), wherein the "formulation of Jordan's and Dirac's ideas," they say, "[becomes] substantially simpler and therefore more transparent and more easily understandable." One presumes that it is this simplicity, transparency, and understandability that makes the analytic apparatus, as they said at the outset, "well-situated" and "capable of no modification" in its capacity as pure mathematics (Von Neumann, 1963, 106). That is, the apparatus they present is *rigorous enough* that they accepted it as broadly correct. Yet it was not *entirely* without problems. Throughout their paper, Hilbert et al. had used a shortcut to get their operator calculus to display the correct (discrete) behavior when necessary:

> In the foregoing we have proceeded as if all variables would vary in a continuous domain, while physically the cases of interest are just those in which conditions are quantized, and therefore the variables also have to run through discontinuous ranges of values. However, for the moment, our proceeding is throughout quite consistent and comprehensive without these last discontinuous cases since we have expressly introduced improper functions $\delta(x - y)$ whose occurrence has no other meaning than that the corresponding variables are only able to take certain discrete values.

Others, notably Dirac, had done this as well. This was a problem:

> However, from the mathematical standpoint, the way of calculating covered, especially when one tries to treat such questions as the above regarding the statistical weights, is rather unsatisfying, since one is never sure to what extent the operations appearing are really to be permitted.

Indeed, as we glean from this last remark—which refers us to von Neumann's (Von Neumann, 1927a)—it was not *only* a problem of mathematical rigor.

This brings us to the last change wrought by the transformation theory, namely, bringing Born's (Born, 1926a) (Born, 1926b) statistical interpretation of Schrödinger's wave function to new heights of importance through its generalization. Jordan (1927, 811), for instance, apparently drawing on ideas from Pauli (Duncan and Janssen, 2009, 20–1), made it quite explicit that (and how) his formalism was to be interpreted statistically (translation from Duncan and Janssen):

> Pauli considers the following generalization: Let q, β be two Hermitian quantum-mechanical quantities, which for convenience we assume to be continuous. Then there is always a function $\phi(q, \beta)$, such that $|\phi(q_0, \beta_0)|^2 dq$ measures the (conditional) probability that, for a given value β_0 of β, the quantity q has a value in the interval $q_0, q_0 + dq$. Pauli calls this function the probability amplitude.

Further, the transformation theory—and Dirac's and Jordan's thoughts thereupon—seemingly had an influence on Heisenberg's articulation of the uncertainty principle in his (Heisenberg, 1927) (Beller, 1985).[14] The centrality of the statistical interpretation is clear in Hilbert et al., too. They begin the paper as follows (Von Neumann, 1963, 105):

> The basic physical idea of the whole theory consists in bringing to light the general probability relations in patches of rigorous functional relationships in ordinary mechanics.
>
> The nature of these relationships is best explained through a particularly important example. If the value W_n of the energy of the system is known, and namely equal to the n-th eigenvalue of the quantized system, then following Pauli the

[14]However, also see fn. 252 of (Mehra and Rechenberg, 2001, 210).

> probability density that the system coordinate has a value between x and $x + dx$ is given by $|\psi_n(x)|^2$, where ψ_n is the eigenfunction associated with the eigenvalue W_n.

But while this understanding begins as an example, it ends as an instance of the general theory (Von Neumann, 1963, 131):

> With this it is recognized that the physical interpretation of the eigenfunctions of Pauli given in the introduction is a special case of the general theory; for now $|\psi_n(x)|^2$ is the probability that the position coordinate of the system has a value between x and $x + dx$ when the system finds itself in the n-th state.

One immediately wonders, however, which states are allowable in this apparatus, i.e., whether there are states that remove this statistical character from the transformation theory. Consequently, the transformation theory quickly led to questions of the completeness (specifically, w.r.t. determinism) of quantum mechanics.

Thus, the situation was this as von Neumann began his own work on quantum mechanics. First, 'quantum mechanics' meant the transformation theory. Given that Schrödinger not only desired but effected an equivalence between matrix and wave mechanics, one presumes he would have accepted the arrival of the transformation theory, as well, at least to some degree. Second, there appeared to be a need for a theory of the Dirac delta and its ilk.[15] Third, Born's statistical interpretation was assumed to interpret the transformation theory, and there immediately came the question of the completeness of quantum mechanics. It was understood that a theory of the Dirac delta might shed light on this question.

[15] Gimeno et al. (2020) argue that, in fact, it was reluctance to use (functions like) the Dirac delta function that doomed Born and Wiener's operator calculus because, without it, they could not solve the problem of linear motion, which was its intended purpose. See (Peters, 2004) for more on the status of (functions like) the Dirac delta at the time. [I recommend the latter reference on recommendation of the former; I have not yet found a copy of this work to review it myself.]

3.4 Von Neumann's Axiomatic Completion of Quantum Mechanics—In 1927

It was von Neumann's aim to answer the completeness question for quantum mechanics using the axiomatic method. Two problems stood in his way, however. First, a theory of the Dirac delta apparently still lay in the way of a definitive answer to the completeness question. Second, it was not clear what, precisely, was being assumed in quantum mechanics. I discuss these in turn, giving only brief attention to the former. However, as an interlude, I introduce a debate between von Neumann and Schrödinger that I place sometime in 1927 or early 1928.

First, the transformation theory was still plagued by the "unrigorous" Dirac delta. More immediately, the occurrence of the Dirac delta in the transformation theory meant that the latter was not yet fully formed mathematically. This von Neumann meant to tackle in (Von Neumann, 1927a)(Von Neumann, 1963, 153):

> [In the transformation theory] [i]t is impossible to avoid including the improper eigenfunctions (see §IX); such as, e.g., $\delta(x)$ first used by Dirac, which is supposed to have the following (absurd) properties: $\delta\left(x\right)=0$,for $x\neq 0$,
> $\int_{-\infty}^{\infty}\delta\left(x\right)dx=1$.
>
> ... But a common deficiency of all these methods is that they introduce in-principle unobservable and physically meaningless elements into the calculation[...]. Although the probabilities appearing as final results are invariant, it is unsatisfactory and unclear why the detour through the non-observable and non-invariant is necessary.

> In the present paper we try to give a method to remedy these shortcomings, and, as we believe, to summarize the statistical standpoint in quantum mechanics in a unitary and rigorous way.

Two things matter from this work. First, von Neumann placed quantum mechanics on a rigorous mathematical footing. In so doing, he entirely avoided the Dirac delta function and, hence, showed that it was irrelevant to the completeness question for quantum mechanics. That is, quantum mechanics now had a *unitary* formalism. Second, he understood the core of quantum mechanics to consist in the solving of eigenvalue problems, much like Schrödinger did wave mechanics (in certain moods).

Before discussing von Neumann's "Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik," I want to introduce a little-known debate that occurred between Schrödinger and von Neumann; with it on the table, we can discuss the details of von Neumann's induction. The topic was the completeness of quantum mechanics. The background was this: in early 1927, there was general agreement among the Göttingen theorists and Dirac that quantum mechanics was essentially statistical (Mehra and Rechenberg, 2001, 210–1). Moreover, despite some internal disagreement about whether this reflected something about the quantum domain per se, there was high confidence that the mathematical formalism was faithful to reality. In Jordan, this led to a sharp criticism of Schrödinger's *Abhandlungen zur Wellenmechanik* (Schrödinger, 1927b) for its reliance on "guiding principles" opposed by "the majority of physicists" (cited and translated in Mehra and Rechenberg, 211). While ending with an apology for the "unfriendly tone," Jordan did not back down from his characterization in a follow-up letter to Schrödinger (16 May 1927; translated in Mehra and Rechenberg, 212):

> It seemed to me that I only reproduced your own views by stating that your interpretation stands in harsh contrast to the fundamental assumptions of Bohr. Now

> it is correct that all quantum-mechanical theoreticians—Bohr, Born, Heisenberg, Pauli, Dirac, Wentzel, Oppenheimer, Gordon, von Neumann—are convinced that the fundamental assumptions of Bohr must be upheld without exception. Therefore, I do not believe that I exaggerated when I stated that the majority of physicists take a standpoint different from yours.

Here we see the front end of the divide we later see between the quantum mechanical theoreticians on one hand and the determinists, including Einstein, Schrödinger, Planck, and von Laue, on the other. Note that von Neumann made the list as a quantum mechanical theoretician. In this sense, then, von Neumann is no different from his fellow quantum mechanical theoreticians.

Von Neumann's debate with Schrödinger reveals the extent to which he was thinking physically and, indeed, thinking about extensions of quantum mechanics. Wigner relates the events in his discussion of Bell's inequality (Wigner 1970, 1009):[16]

> The discussion of Von Neumann, most commonly quoted, is that contained in his book [...], Secs. IV.1 and IV.2. As an old friend of Von Neumann, and in order to preserve historical accuracy, the present writer may be permitted the observation that the proof contained in this book was not the one which was principally responsible for Von Neumann's conviction of the inadequacy of hidden variable theories. Rather, Von Neumann often discussed the measurement of the spin

[16] Several facts point to this argument having occurred sometime between von Neumann's arrival in Berlin and the writing of his book. First, von Neumann appears not to have communicated with Schrödinger prior to Berlin, as he asked Weyl to describe his work to Schrödinger in an effort to win the assistantship to Schrödinger (Letter of 27 June, 1927). Besides, von Neumann's communications with Weyl suggest that von Neumann was not sufficiently familiar with quantum theory prior his time in Göttingen. Thus, the argument did not precede his move to Berlin in Summer 1927. Second, the argument seems to have taken place in person: for one, Wigner seems to have intimate knowledge of both sides; for another, there is no record of the discussion anywhere in von Neumann's surviving documents, whereas we would expect one had it taken place in writing after von Neumann's move to the U.S. Finally, the argument is conceptually of a piece with discussions and inquiries we know were happening at the time (Bacciagaluppi and Crull, 2009) (Bacciagaluppi and Valentini, 2009).

component of a spin-$\frac{1}{2}$ particle in various directions. Clearly, the probabilities for the two possible outcomes of a single such measurement can be easily accounted for by hidden variables [...]. However, Von Neumann felt that this is not the case for many consecutive measurements of the spin component in various different directions. The outcome of the first such measurement restricts the range of values which the hidden parameters must have had before that first measurement was undertaken. The restriction will be present also after the measurement so that the probability distribution of the hidden variables characterizing the spin will be different for particles for which the measurement gave a positive result from that of the particles for which the measurement gave a negative result. The range of the hidden variables will be further restricted in the particles for which a second measurement of the spin component, in a different direction, also gave a positive result. A great number of consecutive measurements will select particles the hidden variables of which are all so closely alike that the spin component has, with a high probability, a definite sign in all directions. However, according to quantum mechanical theory, no such state is possible. Schrödinger raised the objection against this argument that the measurement of a spin component in one direction, while possibly specifying some hidden variables, may restore a random distribution of some other hidden variables. It is this writer's impression that Von Neumann did not accept Schrödinger's objection. His point was that the objection presupposed hidden variables in the apparatus used for the measurement. Von Neumann's argument needs to assume only two apparata, with perpendicular magnetic fields, and a succession of measurements alternating between the two apparata. Eventually, even the hidden variables of both apparata will be fixed by the outcomes of many subsequent measurements of the spin component in their respective directions so that the whole system's

> hidden variables will be fixed. Von Neumann did not publish this apparent refutation of Schrödinger's objection.

Thus, von Neumann was convinced by a *physical* argument that quantum mechanics could not be extended with hidden variables.

Obviously, non-trivial assumptions are being made here, both by von Neumann and by Schrödinger. For ease of understanding, we can put these in our own terms. First, the hidden variables von Neumann has in mind are such that their measurement reveals a property the system already had; this is clear as Wigner says the outcome of a subsequent measurement "restricts the range of values which the hidden parameters must have had before that first measurement was undertaken." Second, the purpose of a would-be hidden variable theory, according to von Neumann, is to "complete" the state description of a system so as to return determinism; we gather this from the general setup, where the goal is to narrow in on initial conditions that can predict future measurements. (In our terms, we only care about effectively-measurable, non-contextual quantities.) Of course, if one is taking quantum mechanics' probability relations for granted—call this assumption (*Uncertainty*)—this is not possible in combination with the first assumption. A detailed version of this line of thinking echoes Heisenberg re: the indeterminacy relations (Beller, 1985, 346–8). Now we come to Schrödinger's objection and von Neumann's reply. From our vantage point, we see Schrödinger as gesturing toward the contextuality response, i.e., that while measurement may reveal aspects of the system, the apparatus itself influences measurement, too. To a Bohmian, this therefore looks like just the right response.

It is tempting to read von Neumann as begging the question against Schrödinger, even as winning him over with rhetoric rather than substance. I think this is too quick, however, and ultimately the episode deserves more scrutiny than I can manage here. Nevertheless, I will make a few observations. First, Schrödinger was already doubting the prospects of his interpretation in late 1926 in light of recent experimental work, so it would not have been von Neumann alone who convinced him; relatedly, it is clear from his subsequent writings that Schrödinger was not one to back down in the face of social pressure. Second, Schrödinger's (and de Broglie's) hope had been for a genuine *matter wave* theory, not a pilot-wave-in-configuration-space theory. Indeed, de Broglie abandoned his approach when it became clear to him this goal was unattainable. Finally, Schrödinger himself sometimes appeared to agree with the quantum mechanical theoreticians—for instance in his 5 May, 1928, letter to Bohr (translation from (Bohr, 1985, 46–8)):

> One further remark: If you want to describe a system, e.g., a mass point by specifying its p and q, then you find that this description is only possible with a limited degree of accuracy. This seems to me very interesting as a limitation in the applicability of the old concepts of experience. But it seems to me imperative to demand the introduction of *new* concepts, with respect to which this limitation *no longer* applies.

So far so good for a Bohmian (modulo the final remark implying he meant to circumvent the uncertainty relations altogether). Except he continues:

> Because what is in principle unobservable should not at all be contained in our conceptual scheme, it should not be possible to represent it within the latter. In the *adequate* conceptual scheme it should no longer appear as if our possibilities of experience were limited through unfavorable circumstances.

Putting aside Schrödinger's reaction to this disagreement for the moment, what this event demonstrates is that von Neumann, among others, believed that quantum mechanics was incompatible with hidden variables on more-or-less intuitive grounds. However, it remained unclear whether the statistical interpretation derived *from* the transformation theory or was merely *assumed* of it. In this way, the situation was analogous to the status of Archimedes' Axiom prior to Hilbert's axiomatization of geometry.

This brings us to the second problem von Neumann faced in 1927. On the Hilbertian understanding, axiomatic theories are given relative to an agreed upon set of propositions, and they are constructed with the help of concepts that set out some or other propositions as fundamental. When Hilbert et al. began the work we find in their joint paper, the scope of quantum theories and especially the concepts involved were unclear. However, with the transformation theory and its statistical interpretation at hand, quantum theorizing was finding its footing on quantum mechanics. Despite this, the foundational concepts remained underspecified, particularly the status of the statistical interpretation. Indeed, as far as von Neumann was concerned, the statistical interpretation had merely been assumed (Von Neumann, 1927c)(Von Neumann, 1963, 209):[17]

> The method commonly used in statistical quantum mechanics was essentially deductive: the absolute square of certain expansion coefficients of the wave function, or of the wave function itself, was equated quite dogmatically with probability, and agreement with experience was subsequently verified. However, a systematic derivation of quantum mechanics from facts of experience or basic assumptions of probability theory, i.e., an inductive foundation, was not given. Also the relation to ordinary probability was an insufficiently clarified one: the validity of its basic laws (addition and multiplication law of probability) was not sufficiently discussed.

Thus, to effect an axiomatic completion, von Neumann yet needed to identify the basic assumptions that give rise to quantum mechanics, i.e., his Hilbert space formulation of it. That is, what did everyone agree *on*?

[17] This paper of von Neumann's was submitted to the *Proceedings* by Born about a month after the 5th Solvay Conference.

> In the present work such an inductive structure is to be attempted. We make the assumption of the unconditional validity of ordinary probability theory. It turns out that this is not only compatible with quantum mechanics, but also (in combination with less far-reaching factual and formal assumptions—compare the summary in §IX, 1-3) sufficient for its unambiguous derivation. Indeed, we will be able to establish the entire 'time-independent' quantum mechanics on this basis.

Note that the two features of the axiomatic method I highlighted earlier are present here. First, the induction is transparently *provisional*—"to be attempted" is weaker even than the more typical academic-ese of "given." This should not be a surprise, either, for in this context *any* attempt is an improvement upon merely assuming the statistical interpretation. Second, he is transparent about the assumptions that he is making. More importantly, he is providing an inductive structure *for* quantum mechanics—that is, he intends to get back quantum mechanics with whatever assumptions he lands on. These assumptions are made *relative* to quantum mechanics. In addition, he makes it clear that his analysis is relative to—that is, uses—the Hilbert space formalism. These are cardinal sins for Bell-style axiomatics; hence, he is working toward an axiomatic completion, not providing an axiomatic reconsideration.

Not surprisingly, the assumptions he makes end up looking much like those he assumed in his disagreement with Schrödinger. Summarizing the (non-probabilistic) assumptions qualitatively, he says in closing (Von Neumann, 1963, 234):

> The goal of the preceding work was to show that quantum mechanics is not only compatible with ordinary probability theory, but rather that under its presupposition—and some plausible factual assumptions—even the only possible solution. The underlying assumptions were the following:

1. Each measurement changes the measured object, and therefore two measurements always interfere with each other—unless one can replace both with one.
2. However, the change caused by one measurement is such that the measurement remains valid, i.e., if you repeat it immediately afterwards, you will find the same result.
3. The physical quantities are—in following a few simple formal rules—to be written as functional operators.

He quickly follows these with: "Note, by the way, that the statistical, "acausal" nature of quantum mechanics is due solely to the (principal!) inadequacy of measurement (cf. the work of Heisenberg cited in notes 2 and 4)." Thus, von Neumann is assuming (*Uncertainty*) Heisenberg's understanding of the uncertainty relations in 1 and 2;[18] (*Quantities*) the quantum mechanical way of representing quantities, which restricts one to just those that are effectively measurable; and (*Probability*) the ordinary probability theory.

We should characterize (*Quantities*) and (*Probability*) further. These assumptions show up in §II, "basic assumptions." Let $\{\mathfrak{S}_1, \mathfrak{S}_2, \mathfrak{S}_3, ...\}$ be an ensemble of copies of the system $\mathfrak{S}$. Given that the goal is to recover quantum mechanics, von Neumann aimed for an expression of the expectation value $Exp(\mathfrak{R})$ in the ensemble of some quantity $\mathfrak{R}$ of the system. The assumption (*Probability*) amounted to:

A. *Linearity.* $Exp(\alpha\mathfrak{R} + \beta\mathfrak{S} +) = \alpha Exp(\mathfrak{R}) + \beta Exp(\mathfrak{S}) + \cdots$, $(\alpha, \beta$ real).

B. *Positive-definiteness.* If the quantity $\mathfrak{R}$ is always positive, then $Exp(\mathfrak{R}) \geq 0$.

while (*Quantities*) amounted to:

[18] In the introduction to (Von Neumann, 1927b), von Neumann introduced the same assumptions, saying "1. Corresponds to the explanation given by Heisenberg for the a-causal behavior of quantum physics; 2. expresses that the theory nonetheless gives the appearance of a kind of causality" (Von Neumann, 1963, 236). Translation by Duncan and Janssen (Duncan and Janssen, 2013, 248).

C. *Linearity of operator assignment to quantities.* If the operators $R, S, \ldots$ represent the quantities $\mathfrak{R}, \mathfrak{S}, \ldots$, then $\alpha R, \beta S, \ldots$ represents the quantity $\alpha\mathfrak{R}, \beta\mathfrak{S}, \ldots$.

D. If the operator R represents the quantity $\mathfrak{R}$, then $f(R)$ represents the quantity $f(\mathfrak{R})$.

As (Duncan and Janssen, 2013, 213) note, assumptions A. and B. do not show up in §IX. Rightly, I am suggesting, they presume this is because they are "part of ordinary probability theory." Indeed, they also note that assumptions 1 and 2 (above quote) do not appear in A.–D.; this is because these assumptions are captured through their correspondence with operators (1 through commutative properties of operators and 2 through idempotency of projection operators corresponding to the measurement being made). Further, von Neumann defines dispersion-free and pure states:[19]

α An $Exp(\mathfrak{R})$ is *dispersion-free* if $Exp(\mathfrak{R}^2) = Exp(\mathfrak{R})^2$

β An $Exp(R)$ is *pure* if $Exp(\mathfrak{R}) = \alpha Exp'(\mathfrak{R}) + \beta Exp''(\mathfrak{R}), \alpha, \beta > 0, \alpha + \beta = 1$ implies $Exp(\mathfrak{R}) = Exp'(\mathfrak{R}) = Exp''(\mathfrak{R})$.

Von Neumann's assumptions A., B. and definitions α, β are not exceptional here. A. was a common assumption for probability at the time, and it was well-fitted to the quantum mechanical view. While it is not stated so explicitly in von Mises, whose work von Neumann was familiar with and later cited, expectation values naturally behave linearly in his *Kollektiv* approach. And as von Neumann hastens to add in a footnote, this also held for non-commuting quantities.[20] One might also be worried that α ruled out an important class of hidden variable theories. However, recall from above that the *point* of a hidden variable

[19]Von Neumann does not label these definitions as he does in his book. Nevertheless, I use these labels for ease of referring back to them.

[20]Duncan and Janssen (2013, 247) (rightly) point out that "[w]hile it may be reasonable to impose condition (A) on directly measurable quantities, it is questionable whether this is also reasonable for hidden variables.". However, this misses that von Neumann, by intending only to capture quantum mechanics as it existed, *meant* to treat only directly measurable quantities.

theory was understood, by seemingly all parties, to be the identification of observable variables returning determinism. But these are just those that would give dispersion-free states. Indeed, Schrödinger himself seemed to accept A. and α;[21] however, given other remarks from the time, it is not clear he held them consistently or reflected on their significance deeply.

Thus we see in his debate with Schrödinger, in his mathematical founding, and in his inductive founding that von Neumann is unabashedly assuming quantum mechanics (or in the latter what is the same, the assumptions that give rise to quantum mechanics). In his debate with Schrödinger, we saw expressed the general presumption that quantum mechanics *per se* can contain no hidden variables. Likewise, we have seen that the transformation theory was taken even by von Neumann himself to be the more-or-less final form of quantum mechanics—though with the caveat that the lack of a theory of the Dirac delta function mucked up a proper understanding of its kinematic completeness. In his mathematical foundations, von Neumann showed that such a theory of the Dirac delta was unnecessary, for the transformation theory could be reinvented from the ground up without reference to it. Indeed, what von Neumann showed was that exactly what one needed to solve the central problem of quantum mechanics—the eigenvalue problem, around which Schrödinger's wave mechanics was built—was the Hilbert space formalism. The sole remaining question, then, was whether von Neumann's Hilbert space formalism *truly* captured quantum mechanics by ruling out hidden variables: if the Hilbert space formalism is the unique representation of quantum mechanics, then hidden variables should be impossible.

[21] For instance, Schrödinger strongly endorsed von Neumann's "Beweis des Ergodensatzes und des H-Theorems in der neuen Mechanik" (von Neumann, 1929) in a December 1929 letter to the latter, saying "The idea of linking the actual operators of quantum mechanics with real measurements contained a dissonance which has now been fundamentally resolved. By using these new concepts extensively it will be possible to achieve a real mapping of the real measurements to the scheme of quantum mechanics, and then the same scheme will be satisfying" (AHQP, Dec 1929 letter). This work explicitly assumed Heisenberg's view of the uncertainty relations and the so-called quantum principle, that quantities correspond to Hermitian operators on a Hilbert space. These are just (*Quantity*) and (*Uncertainty*), and presumably (*Probability*) was also assumed. Additionally, Schrödinger effectively accepted this as late as 1935, where he assumed would-be hidden states are dispersion-free and respect the functional relations between Hermitian operators. See (Bacciagaluppi and Crull, 2009).

There things stood in 1927, and it was not until 1932 that there came a proof to this effect. Except, this matter was essentially settled in 1927—all of the required tools were there! The essential ingredients in his later proof are the trace formula $Exp(R) = Tr(UR)$ and the definition α of dispersion-free expectation values; once those are in place, the proof is trivial, involving some quick definition-chasing. Anyone reading the paper should have seen this, and it is highly unlikely that von Neumann did not see this himself. So why no proof until 1932?

3.5 A Textbook for Mathematicians

It is my contention that the proof did not show up until 1932 because there was no need to publish it until then. An important fact in this regard is that von Neumann's book was part of Courant's series of textbooks for lay-mathematicians, *Basic Teachings of Mathematical Science* (subtitled *in Stand-Alone Presentations with Special Consideration for the Fields of Application*). While von Neumann's book's occurrence as part this series is naturally prominent in the front-matter of the German (Springer) publication, this fact is entirely obscured in subsequent printings. Nonetheless, this tells us that the audience was not expected to be familiar with the area, meaning that results needed to be especially explicit. Meanwhile, the 1927 works were published in the *Göttingen Nachricten* whose readers could be expected to be reasonably familiar with the techniques von Neumann borrowed from the likes of Schmidt, Courant, Toeplitz, Hellinger, and Hilbert. In the rest of this section, I will briefly characterize the book from the standpoint of the axiomatic method. In so doing, it will be clear that von Neumann was using Hilbert's axiomatic method.

One of the primary aims of von Neumann's book (von Neumann, 1932)(Von Neumann, 1955) was to determine whether the Hilbert space formalism—with the "induction" assumptions serving as its basis—could countenance hidden variables. We see hints of this already at the very beginning of the preface (Von Neumann, 1955, vii):[22]

> The object of this book is to present the new quantum mechanics in a unitary [einheitliche] representation which, so far as it is possible and useful, is mathematically unobjectionable [einwandfreie].[...]Therefore the principal emphasis shall be placed on the general and fundamental questions which have arisen in connection with this theory. In particular, the difficult problems of interpretation, many of which are even now not fully resolved, will be investigated in detail. In this context the relation of quantum mechanics to statistics and to the classical statistical mechanics is of special importance.[23]

Fitting for an axiomatization, von Neumann wants a "unitary" representation that is "mathematically unobjectionable"—that is, it is a *unique* and *formal* (=fact-free) mathematical representation. Further, it is a representation and investigation of *quantum mechanics*, as it then existed, and not the more nebulous idea of a "generic" theory of quantum phenomena. It is also clearly provisional, at least in the sense that it does not claim to resolve every problem of interpretation.

The rest of the preface then focuses predominately on the two issues we have already encountered, namely the Dirac delta and the existence of hidden variables. Firstly, von Neumann foreshadows the irrelevance of the Dirac delta fiction for developing quantum mechanics (Von Neumann, 1955, ix):

[22] Page numbers will refer to the English translation's original 1955 printing.

[23] I have provided the original German words in brackets where I depart from Beyer's translation. While I do not disagree with Beyer's translation, I nevertheless think the terms I use better capture the intended meaning in contemporary (philosophical) English. 'Einheitliche' translated as 'unitary' better emphasizes the singleness implied, while I translate 'einwandfrei' more colloquially as 'unobjectionable' to avoid unintended association with the superficial rigor of later sufferers of Theory T syndrome; at any rate, Beyer translates the latter this way on page ix.

> The method of Dirac, mentioned above, (and this is overlooked today in a great part of quantum mechanical literature, because of the clarity and elegance of the theory) in no way satisfies the requirements of mathematical rigor—not even if these are reduced in a natural and proper fashion to the extent common elsewhere in theoretical physics. For example, the method adheres to the fiction that each self-adjoint operator can be put in diagonal form. In the case of those operators for which this is not actually the case, this requires the introduction of "improper" functions with self-contradictory properties. The insertion of such a mathematical "fiction" is frequently necessary in Dirac's approach, even though the problem at hand is merely one of calculating numerically the result of a clearly defined experiment. There would be no objection here if these concepts, which cannot be incorporated into the present day framework of analysis, were intrinsically necessary for the physical theory. Thus, as Newtonian mechanics first brought about the development of the infinitesimal calculus, which, in its original form, was undoubtedly not self-consistent, so quantum mechanics might suggest a new structure for our "analysis of infinitely many variables"—i.e., the mathematical technique would have to be changed, and not the physical theory. But this is by no means the case. It should rather be pointed out that the quantum mechanical "Transformation theory" can be established in a manner which is just as clear and unitary [einheitliche], but which is also without mathematical objections. It should be emphasized that the correct structure need not consist in a mathematical refinement and explanation of the Dirac method, but rather that it requires a procedure differing from the very beginning, namely, the reliance on the Hilbert theory of operators.

As noted above, this dissolves the lingering questions surrounding the transformation theory, and, at the same time, paves the way for determining uniqueness. Not surprisingly, then, von Neumann secondly addresses the axiomatically-fundamental question of the uniqueness of the mathematical representation of the quantum-mechanical view. This begins by noting the inductive foundation of quantum mechanics (Von Neumann, 1955, ix–x):

> In the analysis of the fundamental questions, it will be shown how the statistical formulas of quantum mechanics can be derived from a few qualitative, basic assumptions.

The connection is then made to the uniqueness question:

> Furthermore, there will be a detailed discussion of the problem as to whether it is possible to trace the statistical character of quantum mechanics to an ambiguity (i.e., incompleteness) in our description of nature. Indeed, such an interpretation would be a natural concomitant of the general principle that each probability statement arises from the incompleteness of our knowledge. This explanation "by hidden parameters," as well as another, related to it, which ascribes the "hidden parameter" to the observer and not to the observed system, has been proposed more than once. However, it will appear that this can scarcely succeed in a satisfactory way, or more precisely, such an explanation is incompatible with certain qualitative fundamental postulates of quantum mechanics.

These two explanations more-or-less directly correspond to Schrödinger's hopes, as captured in Wigner's recollection of the debate with von Neumann: von Neumann showed that "according to quantum mechanical theory, no such state [where the spin component has, with a high probability, a definite sign in all directions] is possible"; Schrödinger objected, claiming (essentially) that hidden variables could exist in the measuring apparatus; and von Neumann then showed that the measuring apparatus is no different in kind from the measured system, in the sense that quantum mechanics still applies, hence hidden variables fare no better if posited there. (Again, note that (Probability), (Quantities), and (Uncertainty) are being assumed.) As they occur in the book, these are the arguments of IV.1—2 and VI, respectively.[24]

Before addressing the hidden variables question, von Neumann first recapitulates his earlier work. In chapter I, we receive a summary of the equivalence work that preceded his own, as well as an explanation for its inadequacy for addressing the uniqueness problem and, thereby, for characterizing the "really essential elements of quantum mechanics" (Von Neumann, 1955, 33). This culminates in the following characterization of the goals of chapter II (Von Neumann, 1955, 33):

> We wish then to describe the abstract Hilbert space, and then to prove rigorously the following points:
>
> 1. That the abstract Hilbert space is characterized uniquely by the properties specified, i.e., that it admits of no essentially different realizations.
> 2. That its properties belong to FZ as well as $F\Omega$. (In this case the properties discussed only qualitatively in I.4 will be analyzed rigorously.) When this is accomplished, we shall employ the mathematical equipment thus obtained to shape the structure of quantum mechanics.

Thus in the main, chapter II redescribes von Neumann's work on the Hilbert space formalism, which began with (Von Neumann, 1927a). In chapter III, von Neumann then describes and expands upon the "induction" of quantum mechanics from (Von Neumann, 1927c). In each chapter, especially the latter, it is emphasized throughout that the mathematical formalism is ultimately in service to the quantum mechanical understanding and subject to revision according as the latter itself changes (see, e.g., pp. 133, fn. 86; 211—12; 213—14 with

[24]I only discuss the argument of IV.1—2 here because the argument of VI is also significantly shaped by other contemporaries of von Neumann, particularly Szilard, Bohr, and Heisenberg. I address the latter argument elsewhere.

221—23; 237—38). This is made especially clear in III.2 when von Neumann foreshadows the discussion of hidden variables in IV.1—2 (Von Neumann, 1955, 210):

> Whether or not an explanation of this type, by means of hidden parameters, is possible for quantum mechanics, is a much discussed question. The view that it will sometime be answered in the affirmative has at present prominent representatives. If it were correct, it would brand the present rendering [Form][25] of the theory as provisional, since then the description would be essentially incomplete.
>
> We shall show later (IV.2) that an introduction of hidden parameters is certainly not possible without a basic change in the present theory. For the present, let us re-emphasize only these two things: The ϕ has an entirely different appearance and role from the $q_1, ..., q_k, p_1, ..., p_k$ complex in classical mechanics and the time dependence of ϕ is causal and not statistical: ϕ_{t_0} determines all ϕ_t uniquely, as we saw above.
>
> Until a more precise analysis of the statements of quantum mechanics will enable us to test [prüfen][26] objectively the possibility of the introduction of hidden parameters (which is carried out in the place quoted above), we shall abandon this possible explanation.

Here von Neumann has made it clear (1) that the question is whether *quantum mechanics*—which is mathematically rendered in the Hilbert space formalism—can accommodate hidden variables, and (2) that, contrary to Schrödinger's and others' (e.g., Jordan) expectations, the wave function's evolution is fundamentally unlike that of classical position and momentum in the Hamiltonian schema. In all of this, then, he has made it clear that his axiomatization

[25] I depart from Beyer's translation of 'Form' as 'form' to emphasize that it would be the mathematical form of the theory (quantum mechanics), i.e., the Hilbert space formalism, that is provisional.

[26] Beyer translated 'zu prüfen' as 'to prove', which in typical English implies von Neumann meant to "objectively prove a possibility"; this is certainly not what von Neumann meant, and the more common translation as 'to test' or 'to examine' is more appropriate, regardless.

is *relative* to a set of propositions (quantum mechanics) and to a mathematical formalism (Hilbert space formalism) and *provisional* insofar as quantum mechanics itself is provisional.

3.6 *No* Hidden Variables for Quantum Mechanics

Let us now consider IV.1—2 in this light. By this point, the contents should not surprise us: von Neumann will assume the "inductive" basis of quantum mechanics from his (Von Neumann, 1927c)—the "qualitative basic assumptions"—and examine the possibility of hidden variables. Indeed, this is precisely what happens. I sketch von Neumann's examination in this section.

First, von Neumann makes plain his "basic, qualitative" assumptions. He begins by characterizing the kinds of quantities and relations thereof being considered (Von Neumann, 1955, 297):

> Let us forget the whole of quantum mechanics but retain the following. Suppose a system $\mathfrak{S}$ is given, which is characterized for the experimenter by the enumeration of all the effectively measurable quantities in it and their functional relations with one another. With each quantity we include the directions as to how it is to be measured—and how its value is to be read or calculated from the indicator positions on the measuring instruments. If $\mathfrak{R}$ is a quantity and $f(x)$ any function, then the quantity $f(\mathfrak{R})$ is defined as follows: To measure $f(\mathfrak{R})$, we measure $\mathfrak{R}$ and find the value a (for $\mathfrak{R}$). Then $f(\mathfrak{R})$ has the value $f(a)$. As we see, all quantities $f(\mathfrak{R})$ ($\mathfrak{R}$ fixed, $f(x)$ an arbitrary function) are measured simultaneously with $\mathfrak{R}$. This is a first example of simultaneously measurable quantities. In general, we call two (or more) quantities $\mathfrak{R}$, $\mathfrak{S}$ simultaneously measurable if there is an arrangement which measures both simultaneously in

> the same system—except that their respective values are to be calculated in different ways from the readings. (In classical mechanics, as is well-known, all quantities are simultaneously measurable, but this is not the case in quantum mechanics, as we have seen in III.3.) For such quantities, and a function $f(x, y)$ of two variables, we can also define the quantity $f(\mathfrak{R}, \mathfrak{S})$. This is measured if we measure $\mathfrak{R}, \mathfrak{S}$ simultaneously—if the values a, b are found for these, then the value of $f(\mathfrak{R}, \mathfrak{S})$ is $f(a, b)$. But it should be realized that it is completely meaningless to try to form $f(\mathfrak{R}, \mathfrak{S})$ if $\mathfrak{R}, \mathfrak{S}$ are not simultaneously measurable: there is no way of giving the corresponding measuring arrangement.

Von Neumann then elaborates on non-simultaneously measurable quantities, saying that "their appearance in elementary processes was always to be suspected" and "their presence has now become a certainty" (Von Neumann, 1955, 300–1). Going farther still, he makes it clear that he is taking the uncertainty relations to be general and what are essentially responsible for the intractability of a hidden variable theory (i.e., (*Uncertainty*)). Discussing the attempt to identify hidden variables through successive measurements, von Neumann says that measurement changes the systems such that no progress is made (Von Neumann, 1955, 304–5):

> That is, we do not get ahead: Each step destroys the results of the preceding one, and no further repetition of successive measurements can bring order into this confusion. In the atom we are at the boundary of the physical world, where each measurement is an interference of the same order of magnitude as the object measured, and therefore affects it basically. Thus the uncertainty relations are at the root of these difficulties.

The assumptions, then, are just what were present in von Neumann's (Von Neumann 1927c), namely (*Probability*), (*Uncertainty*), and (*Quantities*). (This last is a bit obscured, but it comes in the fact that the quantities are effectively measurable.) Nothing is new so far, even if the discussion is longer.

Von Neumann then turns to a discussion of hidden variables, beginning by summarizing the intuition one gathers from quantum mechanics (305):

> Therefore we have no method which would make it always possible to resolve further the dispersing ensembles (without a change of their elements) or to penetrate to those homogeneous ensembles which no longer have dispersion. The last ones are the ensembles we are accustomed to consider to be composed of individual particles, all identical, and all determined causally. Nevertheless, we could attempt to maintain the fiction that each dispersing ensemble can be divided into two (or more) parts, different from each other and from it, without a change in its elements. That is, the division would be such that the superposition of two resolved ensembles would again produce the original ensemble. As we see, the attempt to interpret causality as an equality definition led to a question of fact which can and must be answered, and which might conceivably be answered negatively. This is the question: is it really possible to represent each ensemble $[S_1, ..., S_N]$, in which there is a quantity $\mathfrak{R}$ with dispersion, by the superposition of two (or more) ensembles different from one another and from it?

Von Neumann then formalizes the question using the tools of probability theory. In brief, the question is whether there can exist *dispersion-free expectation functions* in quantum mechanics, i.e., whether an ensemble can ever be characterized in a way that all of its variables exhibit no dispersion in the expectation value for their subsequent measurement.

Finally, von Neumann formally characterizes the informal assumptions (*Probability*), (*Uncertainty*), and (*Quantities*) above. Here we should recall that the axiomatic goal is to so precisely define the mathematical formalism that it is the unique characterization of the (informal) theory; if this has been done, then the mathematical formalism should *agree* with any determinations that the (informal) theory makes. In this case, then, a successful axiomatization of quantum mechanics would mean that the Hilbert space formulation agrees with (informal) quantum mechanics that hidden variables are not possible. As in (Von Neumann, 1927c), von Neumann thinks of these assumptions as coming in two types: there are the assumptions of probability and then assumptions specific to quantum mechanics. First, he considers the assumptions of probability, the first several being:

A. If a quantity $\mathfrak{R}$ is always 1, then its expectation is 1, i.e., $Exp(\mathfrak{R}) = 1$;

B. for each $\mathfrak{R}$ and each real number a, $Exp(a\mathfrak{R}) = aExp(\mathfrak{R})$;

C. if $\mathfrak{R}$ is non-negative by nature, then $Exp(\mathfrak{R}) \geq 0$;

D. if the quantities $\mathfrak{R}, \mathfrak{S}, ...$ are simultaneously measurable, then $Exp(\mathfrak{R} + \mathfrak{S} + \cdots) = Exp(\mathfrak{R}) + Exp(\mathfrak{S}) + \cdots$.

A.–C. are obviously trivial, and as von Neumann notes, D. is a theorem of probability. He also notes that it is formulated only for simultaneously measurable $\mathfrak{R}, \mathfrak{S}, \ldots$ "since otherwise $\mathfrak{R} + \mathfrak{S} + \ldots$ is meaningless" (Von Neumann, 1955, 308–9). Yet he continues:

> But the algorithm of quantum mechanics contains still another operation, which goes beyond the one just discussed: namely, the addition of two arbitrary quantities, which are not necessarily simultaneously observable. This operation depends on the fact that for two Hermitian operators, R, S, the sum $R+S$ is also an Hermitian operator, even if the R, S do not commute, while, for example, the product RS is again Hermitian only in the event of commutativity (cf. II.5). In each state ϕ the expectation values behave additively: $(R\phi, \phi) + (S\phi, \phi) = ((R+S)\phi, \phi)$ (cf. E2., III.1). The same holds for several summands. We now incorporate this fact into our general set-up (at this point not yet specialized to quantum mechanics):

> E. if $\mathfrak{R}, \mathfrak{S}, ...$ are arbitrary quantities, then there is an additional quantity $\mathfrak{R} + \mathfrak{S} + \cdots$ (which does not depend on the choice of the $Exp(\mathfrak{R})$-function), such that $Exp(\mathfrak{R} + \mathfrak{S} + \cdots) = Exp(\mathfrak{R}) + Exp(\mathfrak{S}) + \cdots$.
>
> If $\mathfrak{R}, \mathfrak{S}$ are simultaneously measurable, this must be the ordinary sum (by D.). But in general the sum is characterized by E. only in an implicit way, and it shows no way to construct from the measurement directions for $\mathfrak{R}, \mathfrak{S}, \ldots$ such directions for $\mathfrak{R} + \mathfrak{S} + \cdots$.

But this, combined with D., is just A. in (Von Neumann, 1927c). As *many* later commentators have remarked, this is to assume that any would-be hidden variables must behave as if they are quantum mechanical quantities (e.g., (Misra, 1967) (Bell, 1966) (Mermin and Schack, 2018)). One would only assume this if one had already assumed quantum mechanics was true! Yet as I have said, this is exactly right: quantum mechanics—namely, (*Probability*), (*Quantities*), and (*Uncertainty*)—*is* being assumed.

Having formalized the probabilistic aspects of the question, in light of (*Probability*), (*Quantities*), and (*Uncertainty*) just as before,[27] von Neumann then characterizes the relationship quantities will have to the Hilbert space formalism. Yet this, too, is straightforward as this was the entire point of Chapter II and, indeed, von Neumann had already assumed these in the guise of F* and L* in III.5 for his discussion of properties. These correspond to his C. and D. in (Von Neumann, 1927c). Thus, IV.2 begins unremarkably (Von Neumann, 1955, 313–14):

> There corresponds to each physical quantity of a quantum mechanical system, a unique hypermaximal Hermitian operator, as we know (cf., for example, the

[27] Strictly speaking, von Neumann does not use (*Uncertainty*) but a generic indeterminacy relation so as to "not specialize to quantum mechanics." It is important to remember that quantum mechanists believed that Heisenberg's indeterminacy relationship arose in a (nearly) strictly classical fashion, so that *any* physical theory (where energy is discrete) must give rise to a measurement-induced indeterminacy relation. In a sense, this is where the (over)confidence in the generality of Heisenberg's uncertainty relation originates. See, e.g., von Neumann's III.4.

> discussion in III.5.), and it is convenient to assumethat this correspondence is one-to-one—that is, that actually each hypermaximal operator corresponds to a physical quantity. (We also made occasional use of this in III.3.) In such a case the following rules are valid[28] (cf. F., L. in III.5, as well as the discussion at the end of IV.1.):
>
> I. If the quantity $\mathfrak{R}$ has the operator R, then the quantity $f(\mathfrak{R})$ has the operator $f(R)$.
>
> II. If the quantities $\mathfrak{R}, \mathfrak{S}, \ldots$ have the operators $R, S, \ldots$, then the quantity $\mathfrak{R} + \mathfrak{S} + \cdots$ has the operator $R + S + \cdots$. (The simultaneous measurability of $\mathfrak{R}, \mathfrak{S}, \ldots$ is not assumed, cf. the discussion on this point above.)

Likewise, the section ends unremarkably as concerns the status of hidden variables: they cannot be added to the Hilbert space formalism. Unclimactically, then, the uniqueness question has been answered, and in the negative: quantum mechanics (i.e., (*Probability*), (*Quantities*), and (*Uncertainty*)) cannot be extended to include hidden variables, hence insofar as Schrödinger accepted (*Probability*), (*Quantities*), and (*Uncertainty*), he could not have found hidden variables.

Finally, von Neumann turns to discuss the implications of this result. Summarizing the technical meaning, he says (Von Neumann, 1955, 323):

> We have derived all these results from the purely qualitative conditions A'., B'., α), β), I., II.
>
> Hence, within the limits of our conditions, the decision is made and it is against causality; because all ensembles have dispersions, even the homogeneous.

[28]Note also that von Neumann says these rules "are valid" in such a case, rather than that such rules "are true" or something of the sort: he is signaling that (*Quantities*) has already been assumed.

But as we know, A'., B'., α), β), I., II. are just the formalization of (*Probability*), (*Quantities*), and (*Uncertainty*), i.e., quantum mechanics. Rephrasing, then, von Neumann continues (Von Neumann, 1955, 324)(italics mine):

> It should be noted that we need not go any further into the mechanism of the "hidden parameters," since we now know that the established results of quantum mechanics can never be re-derived with their help. In fact, we have even ascertained that it is impossible that the same physical quantities exist with the same function connections (i.e., that I., II. hold), if other variables (i.e., "hidden parameters") should exist in addition to the wave function.
>
> Nor would it help if there existed other, as yet undiscovered, physical quantities, in addition to those represented by the operators in quantum mechanics, because the relations assumed by quantum mechanics (i.e., I., II.) would have to fail already for the by now known quantities, those that we discussed above. *It is therefore not, as is often assumed, a question of re-interpretation of quantum mechanics,–the present system of quantum mechanics would have to be objectively false, in order that another description of the elementary processes than the statistical one be possible.*

This is unambiguously correct. Von Neumann proved that quantum mechanics is uniquely (w.r.t. its kinematical structure) characterized by the Hilbert space formalism, i.e., A'., B'., α), β), I., II.. It follows from this that quantum mechanics does not admit of any re-interpretation. Hence, for another description than the statistical one to be possible—namely, a hidden-variable description—one of quantum mechanics' assumptions must be false. And von Neumann does not consider this impossible (Von Neumann, 1955, 327–8):

> The question of causality could be put to a true test only in the atom, in the elementary processes themselves, and here everything in the present state of our

> knowledge militates against it. The only formal theory existing at the present time which orders and summarizes our experiences in this area in a half-way satisfactory manner, i.e., quantum mechanics, is in compelling logical contradiction with causality. Of course it would be an exaggeration to maintain that causality has thereby been done away with: quantum mechanics has, in its present form, several serious lacunae, and it may even be that it is false, although this latter possibility is highly unlikely, in the face of its startling capacity in the qualitative explanation of general problems, and in the quantitative calculation of special ones. In spite of the fact that quantum mechanics agrees well with experiment, and that it has opened up for us a qualitatively new side of the world, one can never say of the theory that it has been proved by experience, but only that it is the best known summarization of experience.[29]

At the same time that this answers the question of extending quantum mechanics with hidden variables, it also achieves the goals of the axiomatic method. First, von Neumann has ordered the facts of quantum mechanics. It is straightforward how. Second, he oriented our future research. This orientation has not been sufficiently appreciated to date. I briefly discuss this in the conclusion.

3.7 Conclusion: Orienting for Our Future

In conclusion I discuss the goal of the axiomatic method, particularly what it means to *orient* future research. To then orient ourselves, I wish to begin again with Hilbert's "Axiomatische Denken." There we glimpse—however flowery its expression may be—Hilbert's true aim, of enriching mathematics through the sciences and vice versa (Hilbert, 1917, 405):

[29]Note that because the "formal theory"—the Hilbert space theory—was explicitly constructed to be quantum mechanical, and because he has now shown that it is the *unique* such formal theory, the ambiguity here between 'the formal theory of quantum mechanics' and 'quantum mechanics' is justified.

> As in the life of the peoples the individual persons can only prosper when all of the neighboring peoples do well, and as it commands the interest of the states that order prevails not only within each individual state, but also that the relations among the states themselves must be well-ordered, so too is it in the life of the sciences. The significant representatives of mathematical thought, in proper recognition of this, have always demonstrated great interest in the laws and the arrangement in the neighboring sciences and, above all, cultivated the relations to the neighboring sciences, especially to the great kingdoms of physics and epistemology, always to the benefit of mathematics itself. I believe the nature of these relations and the basis of their fruitfulness becomes most plain if I describe to you the one general method of research that appears to be more and more effective in the new mathematics: I mean the axiomatic method.

Not least because von Neumann often said as much himself (Von Neumann, 1954), I think this connection should be minded as we consider what it means to *orient* an area of inquiry.

In the case of von Neumann's axiomatization of quantum mechanics, I think the relationship is this. On the one hand, his axiomatization used the tools of mathematics to tell something to the physicist, namely, that quantum mechanics cannot be extended with hidden variables. This is useful for, as I claim, it changed the places folks looked for hidden variable interpretations and dampened curiosity concerning the physical significance of the Dirac delta. However, at the same time it tells us where we might fruitfully focus attention: (*Probability*), (*Quantities*), and (*Uncertainty*). Indeed, this is what has since taken place, and whenever such attention has borne fruit, von Neumann's proof seems to be mentioned as the inspiration. To pick but one (unexceptional) example (Misra 1967):[30]

[30] I have "translated" Misra's 'quantum mechanics' as 'quantum theory' for the sake of consistency with the foregoing.

> The only justification [of von Neumann's A'., B'., α), β), I., II.] is the a posteriori one that they lead to the usual formalism of quantum [theory]. Such a justification, which is sufficient from an empirical point of view, has little compelling force in the context of the hidden-variable problem. For one is now concerned with the possibility of generalizing the usual formalism of quantum [theory] and the mere fact that a set of postulates leads to the usual formalism cannot be a sufficient recommendation for these postulates.

This is the path of Bohm, de Broglie, Bell, and others, and the essential feature is that *physical* or *epistemological* considerations related to (*Probability*), (*Quantities*), and (*Uncertainty*) predominate. Thus, in a first sense, von Neumann's axiomatic completion of quantum mechanics has *oriented* by focusing our attention on the physical and epistemological considerations that underwrite quantum theory.[31]

However, the axiomatic method is *also* about enriching mathematics. Thus, on the other hand, von Neumann's axiomatization uses physical facts to tell something to the mathematician, namely, that attention should be focused on algebras of non-commuting operators, orthomodular lattices, quantum logics, and the like. This has proven fruitful in mathematics, as von Neumann himself ensured. And it also quickly wrapped back around to physics, where the study of Hilbert spaces and C* algebras gave way, in particular, to sharpenings of von Neumann's "no hidden variables" proof. This is the axiomatic reconsideration path that Misra, Gleason, Jauch, Piron, and others have traveled (Misra, 1967):

> The alternative left to us is to proceed axiomatically in the spirit of von Neumann. Only, one must now start with less stringent postulates than those assumed by VON NEUMANN. The aim of such an axiomatic approach is to isolate the weakest possible assumptions which must be violated for having hidden variables.

[31]And, I might add, he has focused our attention on these considerations in a much more precise way than, say, Bohr did.

> Once such assumptions have been isolated, one can then decide if and how they can be altered so as to allow hidden variables.

What is common on this approach is that *mathematical* considerations related to A'., B'., α), β), I., II. predominate. Thus in a second sense, von Neumann's axiomatization of quantum mechanics has *oriented* by focusing our attention on the mathematical considerations to which quantum theory gives rise.

In the end, then, von Neumann's use of the axiomatic method—his axiomatic completion of quantum mechanics—oriented us toward two distinct but related futures. Just as Hilbert had wanted, von Neumann effectively summarized and clarified where we *had been*—in physics as well as in mathematics—in an effort to identify where we *could go*. The relationship between physics and mathematics, not to mention the fields themselves, has been the better for it.

Concluding Remarks

History has come to occupy a central role in today's philosophy of science. Nevertheless, it is as-yet unclear what methods reliably bring the history of science to bear on today's science or its philosophy. In the foregoing, I have examined two putative entries in the scientific realism debate (Chapters 1 and 2) and one prominent historical problem (Chapter 3). In each, I identified defects of today's philosophy of science that influenced the history told. In Chapter 1, the defect was, at bottom, a predisposition to generality, certainty, and clarity in the analysis of science and its history. In Chapter 2, the defect was abstraction in the absence of a clear goal. In Chapter 3, the defect was the presumption that von Neumann was using axiomatization in the same manner as his commentators.

However, what is more interesting than pointing out defects is where these discussions might go with better methods. In each chapter, I gestured in one or another direction. In Chapter 1, I suggested that the scientific realism debate, as Stein understood it, stood to gain by re-centering on the psychology of scientists at work. As an exemplar of this re-centering, I put forward Nancy Nersessian's cognitive-historical research. In Chapter 2, I suggested that philosophers of science who use history to establish scientific norms would be better-served by engagement with more-concrete problems faced by the sciences. In the case of Stanford, whose primary concern is peer review, the suggestion was that he engage directly with current peer-review reform movements. In Chapter 3, I suggested that the axiomatic method was meant to effect a symbiotic relationship between physics and mathematics.

Here in closing, I point toward two places discussion might go. On the one hand, is the conjunction of mathematics and physics we call axiomatic quantum field theory genuinely symbiotic? On the other hand, might effective quantum field theories be better off in such a symbiotic relationship with mathematics?

These suggestions and their motivating analyses may prove specious. Yet all the same, they appear to engage methods that have proven more reliable than those they replace. In conclusion, I submit that the uses made of science's history in today's philosophy of science deserve further scrutiny.

Bibliography

Bacciagaluppi, Guido and Elise Crull (2009). Heisenberg (and Schrödinger, and Pauli) on hidden variables. *Stud. Hist. Philos. Sci. B Stud. Hist. Philos. Modern Phys.*, 40(4): 374–382.

Bacciagaluppi, Guido and Antony Valentini (2009). *Quantum theory at the crossroads: reconsidering the 1927 Solvay conference*. Cambridge University Press.

Baldwin, John T (2018). *Model theory and the philosophy of mathematical practice*. Cambridge University Press, Cambridge, England.

Becker, Adam (2018). *What is Real?: The Unfinished Quest for the Meaning of Quantum Physics*. Hachette.

Bell, John S (1966). On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38(3): 447–452.

Beller, Mara (1985). Pascual Jordan's influence on the discovery of Heisenberg's indeterminacy principle. *Arch. Hist. Exact Sci.*, 33(4): 337–349.

Bennett, A. and A.L. George (2001). Case studies and process tracing in history and political science: Similar strokes for different foci. In Bennett, A. and C. Elman, editors, *Bridges and Boundaries: Historians, Political Scientists and the Study of International Relations*. MIT Press.

Bhakthavatsalam, Sindhuja and Ian James Kidd (2019). Science, realism, and unconceived alternatives: introduction to the special issue on unconceived alternatives. *Synthese*, 196(10): 3911–3913.

Bird, Alexander (2018). *Thomas Kuhn*. Metaphysics Research Lab, Stanford University, winter 2018 edition.

Bohr, Niels (1985). *Volume 6: Foundations of Quantum Physics I (1926-1932)*. North Holland.

Bolinska, Agnes and Joseph D. Martin (2020). Negotiating history: Contingency, canonicity, and case studies. *Studies in History and Philosophy of Science Part A*, 80: 37 – 46.

Born, Max (1926a). Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik*, 38: 803–827.

Born, Max (1926b). Zur Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik*, 37: 863–867.

Born, Max, Werner Heisenberg, and Pascual Jordan (1925). Zur Quantenmechanik. *Zeitschrift für Physik*, 34: 858.

Born, M, W Heisenberg, and P Jordan (1926). Zur Quantenmechanik. II. *Eur. Phys. J. A*, 35(8-9): 557–615.

Boyd, Richard (1983). On the current status of the issue of scientific realism. *Erkenntnis*, 19(1-3): 45–90.

Boyd, Richard (1989). What realism implies and what it does not. *Dialectica*, 43(1-2): 5–29.

Brading, Katherine (2014). David Hilbert: Philosophy, epistemology, and the foundations of physics. *Metascience*, 23(1): 97–100.

Brading, Katherine and Thomas Ryckman (2018). Hilbert on General Covariance and Causality. In Rowe, David E, Tilman Sauer, and Scott A Walter, editors, *Beyond Einstein: Perspectives on Geometry, Gravitation, and Cosmology in the Twentieth Century*, pages 67–77. Springer New York, New York, NY.

Brading, Katherine A and Thomas A Ryckman (2008). Hilbert's 'Foundations of Physics': Gravitation and electromagnetism within the axiomatic method. *Stud. Hist. Philos. Sci. B Stud. Hist. Philos. Modern Phys.*, 39(1): 102–153.

Bub, Jeffrey (2010). Von Neumann's 'no hidden variables' proof: a re-appraisal. *Found. Phys.*, 40(9-10): 1333–1340.

Bub, Jeffrey (2011). Is von Neumann's 'no hidden variables' proof silly. *Deep Beauty: Mathematical Innovation and the Search for Underlying Intelligibility in the Quantum World*, pages 393–408.

Cartwright, Nancy (2012). Rcts, evidence and predicting policy effectiveness. In *The Oxford Handbook of Philosophy of Social Science*, pages 298–318. Oxford University Press.

Chen, Elliott D (2020). Newton's early metaphysics of body: Impenetrability, action at a distance, and essential gravity. *Stud. Hist. Philos. Sci. B Stud. Hist. Philos. Modern Phys.*

Corry, Leo (2004). *David Hilbert and the axiomatization of physics (1898–1918): From Grundlagen der Geometrie to Grundlagen der Physik*, volume 10. Springer Science & Business Media.

Cushing, James T (1994). *Quantum mechanics: historical contingency and the Copenhagen hegemony*. University of Chicago Press.

Dawisha, Adeed, John Ehrenberg, Bruce Gilley, Stephen M. Walt, and Elizabeth Saunders (2013). Review: Ideology, realpolitik, and us foreign policy: A discussion of frank p. harvey's "explaining the iraq war: Counterfactual theory, logic and evidence". *Perspectives on Politics*, 11(2): 578–592.

Detlefsen, Michael (2014). Completeness and the ends of axiomatization. In Kennedy, Juliette, editor, *Interpreting Gödel*, pages 59–77. Cambridge University Press, Cambridge.

Dieks, Dennis (2017). Von Neumann's impossibility proof: Mathematics in the service of rhetorics. *Stud. Hist. Philos. Sci. B Stud. Hist. Philos. Modern Phys.*, 60: 136–148.

Dirac, P A M (1927). The physical interpretation of the quantum dynamics. *Proc. R. Soc. Lond. A Math. Phys. Sci.*, 113(765): 621–641.

Duncan, Anthony and Michel Janssen (2009). From canonical transformations to transformation theory, 1926–1927: The road to Jordan's Neue Begründung. *Stud. Hist. Philos. Sci. B Stud. Hist. Philos. Modern Phys.*, 40(4): 352–362.

Duncan, A and M Janssen (2013). (Never) Mind your p's and q's: Von Neumann versus Jordan on the foundations of quantum theory. *Eur. Phys. J. H*, 38(2): 175–259.

Eder, Günther and Georg Schiemer (2018). Hilbert, duality, and the geometrical roots of model theory. *Rev. Symb. Log.*, 11(1): 48–86.

Einstein, Albert (1921). Geometrie und Erfahrung. In Einstein, Albert, editor, *Geometrie und Erfahrung: Erweiterte Fassung des Festvortrages Gehalten an der Preussischen Akademie der Wissenschaften zu Berlin am 27. Januar 1921*, pages 2–20. Springer Berlin Heidelberg, Berlin, Heidelberg.

Fahrbach, Ludwig (2011). How the growth of science ends theory change. *Synthese*, 180(2): 139–155.

Farrell, Margaret (2020). What would imaginary ancestors do? Unpublished draft.

Frege, Gottlob, Gottfried Gabriel, Brian McGuinness, and Hans Kaal (1980). *Philosophical and mathematical correspondence.* Oxford: Basil Blackwell.

Gimeno, Gonzalo, Mercedes Xipell, and Marià Baig (2020). Operator calculus: the lost formulation of quantum mechanics. *Arch. Hist. Exact Sci.*

Giovannini, Eduardo N (2016). Bridging the gap between analytic and synthetic geometry: Hilbert's axiomatic approach. *Synthese*, 193(1): 31–70.

Glymour, Clark, David Danks, Bruce Glymour, Frederick Eberhardt, Joseph Ramsey, Richard Scheines, Peter Spirtes, Choh Man Teng, and Jiji Zhang (2010). Actual causation: a stone soup essay. *Synthese*, 175(2): 169–192.

Hallett, Michael (1990). Physicalism, reductionism & Hilbert. In Irvine, A D, editor, *Physicalism in mathematics*, pages 183–257. Springer.

Hallett, Michael (1994). Hilbert's Axiomatic Method and the Laws of Thought. In George, Alexander, editor, *Mathematics and Mind*, pages 158–200. Oxford University Press, Oxford, England.

Hallett, Michael (2008). Reflections on the Purity of Method in Hilbert's Grundlagen der Geometrie. In *The Philosophy of Mathematical Practice*, pages 198–255. Oxford University Press.

Harvey, Frank P (2011). *Explaining the Iraq War: counterfactual theory, logic and evidence*. Cambridge University Press.

Heisenberg, W (1925). Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Eur. Phys. J. A*, 33(1): 879–893.

Heisenberg, Werner (1926). Quantenmechanik. *Naturwissenschaften*, 14(45): 989–994.

Heisenberg, W (1927). Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Eur. Phys. J. A*, 43(3-4): 172–198.

Hilbert, David (1917). Axiomatisches denken. *Math. Ann.*, 78(1): 405–415.

Hilbert, David, J v Neumann, and Lothar Nordheim (1928). Über die grundlagen der quantenmechanik. *Math. Ann.*, 98(1): 1–30.

Jordan, Pascual (1927). Über eine neue Begründung der Quantenmechanik. *Zeitschrift für Physik*, 40: 809–838.

Kuhn, Thomas S. (1970). *The Structure of Scientific Revolutions*. University of Chicago.

Lacki, Jan (2000). The early axiomatizations of quantum mechanics: Jordan, von Neumann and the continuation of Hilbert's program. *Arch. Hist. Exact Sci.*, 54(4): 279–318.

Laudan, Larry (1978). *Progress and its problems: Towards a theory of scientific growth*, volume 282. Univ of California Press.

Laudan, Larry (1981). A confutation of convergent realism. *Philosophy of Science*, 48(1): 19–49.

Levy, Jack S (2015). Counterfactuals, causal inference, and historical analysis. *Security Studies*, 24(3): 378–402.

Lewis, Peter J (2001). Why the pessimistic induction is a fallacy. *Synthese*, 129(3): 371–380.

Maddy, Penelope (2007). *Second Philosophy: A Naturalistic Method*. Oxford University Press.

Maddy, Penelope (2020). On the question of realism. Unpublished draft.

Magnus, P D and Craig Callender (2004). Realist ennui and the base rate fallacy. *Philosophy of Science*, 71(3): 320–338.

Manne, Kate (2020). *Entitled: How Male Privilege Hurts Women*. Potter/Ten Speed/Harmony/Rodale.

Mehra, J and H Rechenberg (1987). *The historical development of quantum theory / Erwin Schrödinger and the rise of wave mechanics / the creation of wave mechanics: Early response and applications 1925-1926*. The Historical Development of Quantum Theory. Springer, Berlin, Germany.

Mehra, Jagdish and Helmut Rechenberg (2001). *The historical Development of Quantum Theory. The Completion of Quantum Mechanics, 1926-1941*. Springer, New York, NY.

Mermin, N David and Rüdiger Schack (2018). Homer nodded: von Neumann's surprising oversight. *Found. Phys.*, 48(9): 1007–1020.

Misra, B (1967). When can hidden variables be excluded in quantum mechanics? *Nuovo Cimento: C: Geophys. Space Phys.*, 47(4): 841–859.

Mitsch, Chris (2020a). The (not so) hidden contextuality of von neumann's "no hidden variables" proof. Unpublished draft.

Mitsch, Chris (2020b). Philosophy and psychology meet: Comparing realism and happiness. Unpublished draft.

Mizrahi, Moti (2013). The pessimistic induction: a bad argument gone too far. *Synthese*, 190(15): 3209–3226.

Mizrahi, Moti (2015). Historical inductions: new cherries, same old cherry-picking. *International Studies in the Philosophy of Science*, 29(2): 129–148.

Muller, F A (1997a). The equivalence myth of quantum mechanics —Part I. *Stud. Hist. Philos. Sci. B Stud. Hist. Philos. Modern Phys.*, 28(1): 35–61.

Muller, F A (1997b). The equivalence myth of quantum mechanics—part II. *Stud. Hist. Philos. Sci. B Stud. Hist. Philos. Modern Phys.*, 28(2): 219–247.

Muller, F A (1999). The equivalence myth of quntum mechanics (addendum). *Stud. Hist. Philos. Sci. B Stud. Hist. Philos. Modern Phys.*, 30(4): 543–545.

Nersessian, Nancy (1977). Scientific evolutions: On changing conceptual structures in science. Ann Arbor, Michigan: University Microfilms. Howard Stein and Raymond J. Nelson, supervisors.

Nersessian, Nancy J (2010). *Creating scientific concepts*. MIT press.

Osbeck, Lisa M, Nancy J Nersessian, Kareen R Malone, and Wendy C Newstetter (2011). *Science as Psychology: Sense-Making and Identity in Science Practice*. Cambridge University Press.

Park, Seungbae (2011). A confutation of the pessimistic induction. *Journal for General Philosophy of Science*, 42(1): 75–84.

Park, Seungbae (2019). How to formulate scientific realism and antirealism. *Journal for General Philosophy of Science*.

Peckhaus, Volker (2003). The pragmatism of hilbert's programme. *Synthese*, 137(1/2): 141–156.

Perovic, Slobodan (2008). Why were Matrix Mechanics and Wave Mechanics considered equivalent? *Stud. Hist. Philos. Sci. B Stud. Hist. Philos. Modern Phys.*, 39(2): 444–461.

Peters, Klaus-Heinrich (2004). *Schönheit, Exaktheit, Wahrheit (der Zusammenhang von Mathematik und Physik am Beispiel der Geschichte der Distributionen)*. PhD thesis.

Proctor, Robert W and E J Capaldi (2012). *Psychology of Science: Implicit and Explicit Processes*. OUP USA.

Putnam, Hilary (1975). *Philosophical Papers: Mathematics, Matter and Method*, volume 1. Cambridge University Press.

Rédei, Miklós (1996). Why John von Neumann did not like the Hilbert space formalism of quantum mechanics (and what he liked instead). *Studies in History and Philosophy of Modern Physics*, 4(27): 493–510.

Rédei, Miklós (2006). John von Neumann on quantum correlations. In *Physical Theory and its Interpretation*, pages 241–252. Springer.

Rédei, Miklós and Michael Stöltzner (2006). Soft Axiomatisation: John von Neumann on Method and von Neumann's Method in the Physical Sciences. In *The Western Ontario Series in Philosophy of Science*, pages 235–249. Springer-Verlag, Berlin/Heidelberg.

Reid, Thomas (1852). *The Works of Thomas Reid, D.D.* Edinburgh: Maclachlan and Stewart. London: Longman, Brown, Green, and Longmans.

Ross-Hellauer, Tony (2017). What is open peer review? a systematic review. *F1000Res.*, 6: 588.

Rowe, David E (2018). *A Richer Picture of Mathematics: The Göttingen Tradition and Beyond*. Springer.

Saatsi, Juha (2019). Historical inductions, old and new. *Synthese*, 196(10): 3979–3993.

Sauer, Tilman and Ulrich Majer, editors (2009). *David Hilbert's Lectures on the Foundations of Physics 1915-1927: Relativity, Quantum Theory and Epistemology*. Springer, Berlin, Heidelberg.

Schneider, Mike D (2019). Betting on future physics. *Br. J. Philos. Sci.*

Schrödinger, Erwin (1926). Über das Verhältnis der Heisenberg-Born-Jordanschen Quantenmechanik zu der meinem.

Schrödinger, Erwin (1927a). *Abhandlungen zur Wellenmechanik*. Barth, Leipzig.

Schrödinger, Erwin (1927b). *Collected Papers on Wave Mechanics*. Chelsea, New York.

Sieg, Wilfried (2014). The Ways of Hilbert's Axiomatics: Structural and Formal. *Perspect. Sci.*, 22(1): 133–157.

Smith, George (2014). Closing the loop: Testing newtonian gravity, then and now. In *Newton and Empiricism*, pages 31–70. Oxford University Press.

Spelke, ES, E Newport, and R Lerner (1998). Nativism, empiricism, and the development of knowledge. *Handbook of child psychology, 5th ed., Vol. 1: Theoretical models of human development*, pages 275–340.

Stanford, P Kyle (2001). Refusing the devil's bargain: What kind of underdetermination should we take seriously? *Philosophy of Science*, 68(S3): S1–S12.

Stanford, P. Kyle (2005). Instrumentalism. In *The Philosophy of Science*, pages 400–400. Psychology Press.

Stanford, P Kyle (2006). *Exceeding our grasp: Science, history, and the problem of unconceived alternatives*, volume 1. Oxford University Press.

Stanford, P Kyle (2015). Catastrophism, uniformitarianism, and a scientific realism debate that makes a difference. *Philos. Sci.*, 82(5): 867–878.

Stanford, P Kyle (2016). Naturalism without scientism. *The Blackwell companion to naturalism*, pages 91–108.

Stanford, P Kyle (2017). Unconceived alternatives and the strategy of historical ostension. In *The Routledge handbook of scientific realism*, pages 212–224. Routledge.

Stanford, P Kyle (2019). Unconceived alternatives and conservatism in science: the impact of professionalization, peer-review, and big science. *Synthese*, 196(10): 3915–3932.

Stanford, P Kyle (2020). Realism, instrumentalism, particularism: A middle path forward in the scientific realism debate.

Stein, Howard. Newton: Philosophy of inquiry and metaphysics of nature. Unpublished manuscript.

Stein, Howard. Physics and philosophy meet: the strange case of poincaré. Unpublished manuscript.

Stein, Howard. Physics and philosophy meet: The strange case of poincaré. Unpublished.

Stein, Howard (1958). *An Examination of Some Aspects of Natural Science*. PhD thesis, University of Chicago. University of California: Microfilm.

Stein, Howard (1970). On the notion of field in newton, maxwell, and beyond. In Stuewer, Roger B., editor, *Historical and Philosophical Perspectives of Science*, pages 264–287. Minnesota Studies in the Philosophy of Science, vol. V; University of Minnesota Press.

Stein, Howard (1974). Maurice clavelin on galileo's natural philosophy. *British Journal for the Philosophy of Science*, 25: 375–397.

Stein, Howard (1982). On the present state of the philosophy of quantum mathematics. In *PSA 1982*, pages 563–581. vol. 2.

Stein, Howard (1987). After the baltimore lectures: Some philosophical remarks on the subsequent development of physics. In Kargon, Robert and Peter Achinstein, editors, *Kelvin's Baltimore Lectures and Modern Theoretical Physics*, pages 375–398. MIT Press.

Stein, Howard (1988). Logos, logic, and logistiké: Some philosophical remarks on the nineteenth century transformation of mathematics. In *History and philosophy of modern mathematics*, volume 11, pages 238–259. Minnesota studies in the philosophy of science: Minnesota University Press.

Stein, Howard (1989). Yes, but...: Some skeptical reflections on realism and anti-realism. *Dialectica*, 43: 47–65.

Stein, Howard (1990a). 'from the phenomena of motions to the forces of nature': Hypothesis or deduction? In *PSA 1990*, pages 209–222. vol. 2.

Stein, Howard (1990b). On locke, 'the great huygenius, and the incomparable mr. newton'. In Bricker, Phillop and R.I.G. Hughes, editors, *Philosophical Perspectives on Newtonian Science*, pages 17–47. MIT Press.

Stein, Howard (1992). Was carnap entirely wrong, after all? *Synthese*, pages 275–295.

Stein, Howard (1993). On philosophy and natural philosophy in the seventeenth century. *Midwest Studies in Philosophy*, 18: 177–201.

Stein, Howard (2004). The enterprise of understanding and the enterprise of knowledge–for isaac levi's seventieth birthday. *Synthese*, 140: 135–176.

Stöltzner, Michael (1999). What John von Neumann Thought of the Bohm Interpretation. In Greenberger, Daniel, Wolfgang L Reiter, and Anton Zeilinger, editors, *Epistemological and Experimental Perspectives on Quantum Physics*, Vienna Circle Institute Yearbook [1999], pages 257–262. Springer Netherlands, Dordrecht.

Stöltzner, Michael (2001). Opportunistic axiomatics — Von Neumann on the methodology of mathematical physics. In *John von Neumann and the Foundations of Quantum Physics*, pages 35–62. Springer Netherlands, Dordrecht.

Sun, P., C. Chubb, C.E. Wright, and G. Sperling (2016). The centroid paradigm: Quantifying feature-based attention. *Attention, Perception, and Psychophysics*, 78(2): 474–515.

Tennant, Jonathan P, Jonathan M Dugan, Daniel Graziotin, Damien C Jacques, François Waldner, Daniel Mietchen, Yehia Elkhatib, Lauren B Collister, Christina K Pikas, Tom Crick, Paola Masuzzo, Anthony Caravaggi, Devin R Berg, Kyle E Niemeyer, Tony Ross-Hellauer, Sara Mannheimer, Lillian Rigling, Daniel S Katz, Bastian Greshake Tzovaras, Josmel Pacheco-Mendoza, Nazeefa Fatima, Marta Poblet, Marios Isaakidis, Dasapta Erwin Irawan, Sébastien Renaut, Christopher R Madan, Lisa Matthias, Jesper Nørgaard Kjær, Daniel Paul O'Donnell, Cameron Neylon, Sarah Kearns, Manojkumar Selvaraju, and Julien Colomb (2017). A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Res.*, 6: 1151.

Tirole, Jean (2017). *Economics for the common good.* Princeton University Press.

Treisman, A. and G. Gelade (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1): 97–136.

Tylén, Kristian, Peer Christensen, Andreas Roepstorff, Torben Lund, Svend Østergaard, and Merlin Donald (2015). Brains striving for coherence: Long-term cumulative plot formation in the default mode network. *NeuroImage*, 121: 106–114.

van der Waerden, B.L. (1967). *Sources of Quantum Mechanics.* North-Holland Publishing Company, Amsterdam.

Von Neumann, John (1927a). Mathematische Begründung der Quantenmechanik. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1927: 1–57.

Von Neumann, John (1927b). Thermodynamik quantenmechanischer Gesamtheiten. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1927: 273–291.

Von Neumann, John (1927c). Wahrscheinlichkeitstheoretischer aufbau der quantenmechanik. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1927: 245–272.

von Neumann, J (1929). Beweis des Ergodensatzes und des H-Theorems in der neuen Mechanik. *Zeitschrift für Physik*, 57(1-2): 30–70.

von Neumann, Johann (1932). *Mathematische Grundlagen der Quantenmechanik*, volume XXXVIII of *Die Grundlehren der Mathematischen Wissenschaften.* Verlag von Julius Springer, Berlin.

Von Neumann, John (1954). The Role of Mathematics in the Sciences and in Society. *Graduate Alumni*, pages 16–29.

Von Neumann, John (1955). *Mathematical Foundations of Quantum Mechanics.* Princeton University Press, Princeton.

Von Neumann, John (1963). *Collected works: Logic, theory of sets, and quantum mechanics v. 1.* Pergamon Press, London, England.

Weatherall, J. O. (2018). Theoretical Equivalence in Physics. *ArXiv e-prints.*

Williams, Porter (2019). Scientific realism made effective. *Br. J. Philos. Sci.*, 70(1): 209–237.

Wilson, Mark (2017). *Physics Avoidance.* Oxford University Press, London, England.

Wilson, Mark (2020). Imitations of Rigor.

Winter, A.N., C. Wright, C. Chubb, and G. Sperling (2016). Conjunctive targets are hard in visual search but easy in centroid judgments. *Journal of Vision*, 16(12): 750–750.

Yan, Karen, Meng-Li Tsai, and Tsung-Ren Huang (2020). Improving the quality of case-based research in the philosophy of contemporary sciences. *Synthese*, pages 1–20.

Yang, XiaoHong, HuiJie Li, Nan Lin, Xiuping Zhang, YinShan Wang, Ying Zhang, Qian Zhang, XiNian Zuo, and YuFang Yang (2019). Uncovering cortical activations of discourse comprehension and their overlaps with common large-scale neural networks. *NeuroImage*, page 116200.

Zhao, Kino (2020). Sample representation in the social sciences. *Synthese*, pages 1–19.