8-11-2019

Naturalized Metaphysics and Scientific Constraint: A Model-Building Approach

Jake Spinella

NATURALIZED METAPHYSICS AND SCIENTIFIC CONSTRAINT: A MODEL-BUILDING APPROACH

by

JAKE SPINELLA

Under the Direction of Dr. Daniel Weiskopf, PhD

ABSTRACT

A problem with recent work about the relationship between metaphysics and science, especially in the theorizing of those who identify as "naturalized metaphysicians", is the spotty, metaphorical characterization of what it means for science to "constrain" metaphysics. The most robust account of scientific constraint on metaphysical theorizing is advanced by James Ladyman and Don Ross in their 2007 book *Every Thing Must Go*. Ladyman & Ross claim that the only legitimate metaphysical hypotheses are those that unify two previously disparate scientific explanations. I will critique Ladyman & Ross' account of naturalized metaphysics (and, by extension, their view of science's constraint on metaphysics), and offer an alternative view of naturalized metaphysics as a practice of constructing *physically possible* models of reality. This account yields a

different view of science's constraint on metaphysics, specifically, that models must be physically possible in order to be of methodological and heuristic use to scientists.

INDEX WORDS: Naturalized metaphysics, Naturalism, Model-building, Inference to the Best Explanation, Bayesianism, Metaphilosophy

NATURALIZED METAPHYSICS AND SCIENTIFIC CONSTRAINT: A MODEL-BUILDING APPROACH

by

JAKE SPINELLA

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2019

2019

NATURALIZED METAPHYSICS AND SCIENTIFIC CONSTRAINT: A MODEL-BUILDING APPROACH

by

JAKE SPINELLA

Committee Chair: Daniel Weiskopf

Committee: Andrea Scarantino

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2019

DEDICATION

For my grandmother, Deidre.

ACKNOWLEDGEMENTS

There are enough people who contributed in some way to the production of this thesis that to thank all of them would be impractical. For the sake of brevity, I will focus on only a few people in particular. First, I want to thank my thesis advisor, Dr. Daniel Weiskopf, and my committee member, Andrea Scarantino for their invaluable comments on numerous manuscript drafts. None of this would have been done without your generous support, and your incisive criticisms. I'd also like to thank my close friends, my brother Phillip, and my partner Magnolia for doing all the things they do to keep me moving in the right direction. To convey adequately how important you all are to me would take longer than the length of this thesis.

TABLE OF CONTENTS

1 INTRODUCTION

A problem with much recent theorizing about the relationship between metaphysics and science, especially in the theorizing of those who identify as "naturalized metaphysicians"[1], is the spotty, metaphorical characterization of what it means for science to "constrain" metaphysics. Most metaphysicians take consistency with science as a desideratum of a metaphysical theory but analyze the meaning of 'consistency' differently and demand it to varying degrees.[2] The most robust account of scientific constraint on metaphysical theorizing is advanced by James Ladyman and Don Ross in their 2007 book *Every Thing Must Go*. Ladyman & Ross (henceforward referred to as L&R) argue that the current state of analytic metaphysics is a sorry one, suspiciously reliant on what they see as intuition-mongering divorced from the empirical-scientific realities that would substantially inform debates on fundamental ontology. On their view, metaphysics is strongly constrained by science insofar as the only useful role metaphysics could serve is the unification of previously disparate scientific hypotheses; any metaphysical theorizing which cannot be pressed into the service of scientific unification should be summarily discarded. I will critique L&R's view of naturalized metaphysics (henceforth referred to as NM) and offer a competing account of how NM should conduct itself. Section I will detail L&R's account of NM, focusing on the argument for their "Principle of Naturalistic Closure" (hereafter referred to as PNC), which demarcates legitimate and illegitimate metaphysical hypotheses by their ability to unify two previously disparate specific

[1] "Naturalized Metaphysics" or "Scientific Metaphysics" is a term that is still in search of a rigorous definition by those who fly its banner. Definitions abound, but, in the attempt to be as neutral as possible, I want to describe a naturalized metaphysician as someone who thinks that metaphysical inquiry is constrained by science in a stronger sense than traditional metaphysicians who demand only that metaphysics not contradict the results of our best science. This is deliberately vague, capturing as it does the plurality of different ideas of what it means for metaphysics to be robustly constrained by science.
[2] See Sider (2001), Ladyman & Ross (2007), Maddy (2007), Maudlin (2007), Paul (2012), and Bryant (2017), for evidence of the claim that metaphysicians commonly endorse scientific constraints on the practice of metaphysics.

scientific hypotheses. Section II will give two reasons for why their justification of the PNC actually undermines the prospects of this view of NM. Section III will address the issues raised by L&R's deference to scientific practice, namely the entailed commitment to the truth-conduciveness of a naturalized version of inference to the best explanation (henceforth referred to as IBE) in science. Section IV will respond to an objection that takes the form of an analogy between L&R's naturalized IBE and Bayesian conditionalization: If we take subjective Bayesian conditionalization to be legitimate for theory confirmation, why is naturalized IBE's subjectivity a problem for theory choice? Section V will explore how L&R's view of NM can be salvaged by reconceptualizing the project of NM as one of building physically possible models of reality. This shift of focus to physically possible model-building also yields a different account of how science constrains metaphysical practice.

2 LADYMAN & ROSS' ARGUMENT FOR UNIFICAITON

For L&R, legitimate metaphysical theorizing must obey (along with the "Primacy of Physics Constraint", which won't be addressed in this paper)[3] what they call the "Principle of Naturalistic Closure":

> Any new metaphysical claim that is to be taken seriously at time *t* should be motivated by, and only by, the service it would perform, if true, in showing how two or more specific scientific hypotheses, at least one of which is drawn from fundamental physics, jointly explain more than the sum of what is explained by the two hypotheses taken separately.[4]

[3] The Primacy of Physics Constraint (PPC) states: "Special science hypotheses that conflict with fundamental physics, or such consensus as there is in fundamental physics, should be rejected for that reason alone. Fundamental physical hypotheses are not symmetrically hostage to the conclusions of the special sciences." Ladyman & Ross (2007) Pg. 37.

[4] Ibid. Pg. 30.

L&R's justification for the PNC is, simply, that working scientists take unification seriously, and so it should be the standard by which metaphysical claims are judged:

> Why should radical methodological naturalists suppose that there is any 'responsible and significant' job for metaphysics to do? Our answer is that one of the important things we want from science is a relatively unified picture of the world. We do not assert this as a primitive norm. Rather…it is exemplified in the actual history of science. Scientists are reluctant to pose or to accept hypotheses that are stranded from the otherwise connected body of scientific beliefs. This is rational, reflecting the fact that a stranded hypothesis represents a mystery, and therefore calls out for scientific work aimed at eliminating it. It also reflects the fact that an important source of justification for a hypothesis is its standing in reciprocal explanatory relationships—networked consilience relationships—with other hypotheses.[5]

The fact that working scientists take explanatory unification seriously as a theoretical virtue of scientific practice is what justifies, for L&R, the privileging of it as a norm for naturalized metaphysical inquiry. Theoretical virtues are desiderata like simplicity, explanatory power, predictive power, consilience, etc. According to L&R, unification is the overriding theoretical virtue in scientific inquiry. Since scientists take unification very seriously and naturalized metaphysicians ought to model their practices on the sciences to the best of their ability, a unificationist and naturalist justification of metaphysical explanation is offered in the form of the Principle of Naturalistic Closure. What is the motivation for the normative claim that NM ought to model its practices on the sciences? L&R's response to this query is that science is the sole reliable method for coming to objective knowledge of the natural world:

[5] Ibid. Pg. 28.

> Since science just *is* our set of institutional error filters for the job of discovering the objective character of the world—that and no more but also that *and no less*— science respects no domain restrictions and will admit no epistemological rivals (such as natural theology or purely speculative metaphysics). With respect to anything that is a putative fact about the world, scientific institutional processes are absolutely and exclusively authoritative.[6]

On this scientistic view[7], in order for metaphysics to be truth-conducive it must by necessity model its practices on the sciences, which is why L&R are highly deferential to scientific practice—there simply is no other game in town. In the following section, I will examine why this deference to the practices of working scientists undermines L&R's account of NM.

3 THE DANGERS OF DEFERENCE

L&R's deferring to scientific practice may seem like an intuitive move for a naturalized metaphysician to make: if being a naturalist involves taking science seriously, surely part of taking science seriously involves a high—if not overriding—degree of deference on the part of philosophy to scientific practice. But this deference causes problems for L&R: First, if it is the case that we should privilege those theoretical virtues that scientists take as relevant for theory choice, no compelling evidence has been offered for why scientific unification, rather than any other theoretical virtue or some combination of them, should be the standard by which metaphysical explanations are judged. It is true that unification has been valued throughout the history of science and it is obvious that contemporary working scientists do take unification seriously when comparing competing hypotheses. But it has not been made clear that scientists historically and currently take it to be the supreme theoretical virtue. Scientists then and now also take seriously predictive power, explanatory novelty, explanatory depth, ontological and mathematical simplicity, aesthetic criteria like "elegance"

[6] Ibid. Emphasis theirs.
[7] The chapter the above quote is cited from is titled "In Defence of Scientism".

or "beauty", and so on. For L&R's justification of the PNC to succeed, they need to clarify why scientists take the theoretical virtue of unification more seriously than any other theoretical virtue to the point where any theory that unifies best wins out over every competitor.

No explicit case is made for unification as the supreme theoretical virtue of working scientists in *Every Thing Must Go*, and I don't think it's a plausible case to make in general. Different scientists will value unification, explanatory depth, simplicity, etc. to different degrees and will weigh their sets of preferences idiosyncratically. That scientists historically takes unification seriously is not enough to establish that it's the only thing they take seriously.[8] Since unification is not obviously special—just one virtue among many that we use in judging the plausibility of scientific and metaphysical hypotheses—L&R's motivation for the PNC is vitiated, because scientific practice doesn't really justify the claims that they make on its behalf. *Prima facie*, if we take deference to the theory choice preferences of working scientists seriously, it seems that we should let in all the theoretical virtues that science takes seriously as relevant to our assessment of a particular theory. But Ladyman & Ross cannot simply expand the list of virtues to include all the ones scientists actually value, as doing that is tantamount to abandoning any conception of constraint on metaphysics by science, since metaphysical hypotheses are constrained only by the PNC, and the PNC is not well-motivated. Furthermore, expanding the list of theoretical virtues opens the space for metaphysicians to claim that their theories satisfy these other virtues even if they don't unify two scientific explanations, defanging L&R's critique and calling into question the very possibility of NM as distinct from standard metaphysical practice. Because Ladyman & Ross' strategy is to enforce constraints on metaphysical practice by declaring only a very restricted subset of theoretical virtues

[8] It may be further argued that unification is only an important theoretical virtue in *some* sciences, namely physics and chemistry. Other sciences seem to tolerate (or even encourage) much more disunity than a unificationist could stomach.

(namely, scientific unification) as relevant to theory choice, expanding the list of relevant theoretical virtues is not a live option for them, if they are intent on offering a principled account of science's strong constraint on metaphysical practice, which they are.

Perhaps L&R can respond by shifting the focus of their view from the theoretical virtues to the comparative reliability of scientific method vis a vis metaphysical method, and so defend their picture of scientific constraint by appealing to the reliability of scientific explanation vis a vis empirically ungrounded metaphysics. In a later article, Ladyman does just this, and defends the truth-conduciveness of IBE in science while questioning its truth-conduciveness in metaphysics. The next section will critique Ladyman's argument and make clear why NM cannot be conducted merely by adopting a naturalist criterion of theory choice for IBE.

4 CRITIQUING NATURALIZED IBE

Before evaluating Ladyman's defense of scientific IBE (and, by extension, IBE in NM), a definition of IBE is in order. Following Douven (2002), we can give IBE the following formal definition:

> IBE: Given evidence E and potential explanations $H_1 \ldots H_n$ of E, infer to the (probable/approximate) truth of the H_i that explains E best.[9]

"Explains best" is a phrase that needs unpacking. One standard account of what it is for a theory to "explain the best" is that it possesses the most theoretical virtues vis a vis competing theories.[10] Whatever theory "scores" best in terms of satisficing a weighted plurality of the theoretical virtues is the theory that we ought to infer based on evidence E.

---

[9] Douven (2002). Pg. 356.

[10] See Thagard (1978) and McMullin (1996).

In his 2012 article "Science, Metaphysics, and Method", Ladyman claims that there is a legitimate distinction between IBE as it is used in science and IBE as it is used in metaphysics. What is supposed to distinguish science's reliance on IBE from the analytic metaphysician's is that scientific IBE operates within a tradition where it has a track record of inductive success. Scientific IBEs, like Darwin's formulation of the principle of natural selection as an independent alternative to the theory of independent divine creation, are legitimate because they are verifiable through empirical inquiry and have indeed been verified. IBE in metaphysics, on the other hand, is untethered from any form of worldly epistemic constraint. There are no empirical verification conditions for, say, David Lewis' best systems approach to ascertaining the laws of nature by determining which best strike a balance between simplicity in the number of axioms and strength in the amount of information about the world these axioms impart.[11] Metaphysical IBE, then, is illegitimate by Ladyman's lights because there is no connection to empirical success, so appeals to explanatory power are compelling in scientific IBE but ultimately idle in metaphysical theorizing:

> Explanatory power plays the role it does in theory choice because of the relationship between theoretical explanation and the empirical virtues of scientific theories. We have inductive grounds for believing that pursuing simplicity and explanatory power in science will lead to empirical success, but no such grounds where we are dealing with distinctively metaphysical explanations, since the latter is completely decoupled from empirical success.[12]

Because NM is supposed to generate metaphysical theses that are testable by our best current science via the same methods of IBE that scientists use, IBE in NM is taken to be reliable because the explanatory virtues are not decoupled from empirical success in the way it is for

[11] Lewis (1973).
[12] Ladyman (2012). Pg. 46.

metaphysical IBE.[13] There are two problems with this justification of IBE in science and in NM. The first problem comes from the previously made point that the theory-choice preferences of working scientists are heterogeneous. It's not clear that any working scientist has exactly the same list of theoretical virtues, and this is reflected in diversity of thought in the philosophical literature on explanatory virtues and theory choice: Quine (1966) takes parsimony as supreme in theory choice; Lewis (1973), as previously mentioned, evaluates candidate laws of nature on the basis of their simplicity in being axiomatized and the strength of their informational content; Lipton (2004) claims that a theory's ability to unify disparate phenomena, via a specification of a mechanism that explains with precision why the *explanans* is the way it is constitutes a best explanation in every case; and so on. Other accounts, like Keas (2017), propose upwards of twelve different theoretical virtues. As is made clear from even this cursory glance at the literature on explanation and theory choice, there are myriad ways of listing and organizing the theoretical virtues. This diversity of opinion regarding what actually counts as a theoretical virtue is a point against the reliability of IBE in NM and in science. If theoretical virtues are taken to be confirmationally relevant— that is, if every *bona fide* theoretical virtue must raise the degree of confirmation of a hypothesis that possesses that virtue to a greater degree than all of its theoretical competitors (and, correspondingly, that the degree of confirmation of losing theories is lowered by not possessing that virtue)—then leaving confirmationally relevant virtues out or adding spurious virtues in that are taken to be confirmationally relevant but in fact are confirmationally idle will affect what theory one chooses. So even if we can agree on a particular list of theoretical virtues, there is still a further question about how these virtues are to be weighed against one another; especially when the virtues (like, say, predictive power and explanatory depth) stand in tension with one another, which they often do. If

[13] The requirement that metaphysical theses be in-principle testable is entailed by L&R's endorsement of a "non-positivist" version of verificationism. See *ETMG*, pg. 29 for more.

the hypothesis with the right arrangement of theoretical virtues is most likely to be true, then each theoretical virtue we take into consideration must be confirmationally relevant and must be correctly assigned a weight relative to the rest of the theoretical virtues for us to accurately determine which hypothesis is most likely to be true. Without both of these conditions satisfied, we cannot expect that our answers to the question "Which theory *T* explains evidence *E* the best?" will track what is actually the case. And, given the heterogeneity found in working scientists' listing and weighing of the theoretical virtues, it appears that these two conditions aren't satisfied, so Ladyman cannot explain IBE's success in the sciences by claiming that one only needs a partial list of confirmationally relevant virtues, and each scientist possesses a partial list. And this endangers the efficacy of IBE in science and NM.

The upshot is that without an antecedently known and objective hierarchy of theoretical virtues, we cannot safely assume that we are making the right IBE. Even if we agreed upon a particular list of theoretical virtues, if we value simplicity more than is warranted, or undervalue unification, etc. we may very well choose a theory that is more likely to be false over a theory that is more likely to be true if our agreed-upon listing turns out to be wrong. Therefore, any defender of IBE's efficacy in science or in NM must answer the following two questions: What are the *actual* theoretical virtues, and, after we've decided what they are, how do we go about weighing them? Ladyman is silent on this question.

These questions raise yet another issue, which is: How could we know when we had the right organization of theoretical virtues? Because the theoretical virtues are various, and their criteria of weighting and application are not easily ascertained, it seems that even if everyone were to agree a definitive list of theoretical virtues and a weighting of them, we'd still be left with the question of whether we were right and why. The proper listing of theoretical virtues is, presumably, not something that can be determined *a priori,* but if we try to look to the history of science for guidance on what theoretical virtues reliably lead to the best theories, we quickly find that it is littered with theories that were simple, explanatorily deep, unificatory, and false. If the goal of scientific IBE isn't to increase our understanding but rather to choose the *best* theory—the theory that is *most likely* to be true—it isn't enough to have agreement that there are certain theoretical virtues and that they are important. And without a clear weighting scheme of virtues and an epistemic situation in which we know that our weighting scheme is correct, it seems as if we don't seem to have an answer to the crucial question of "Which theory T explains evidence E the best?", which means that we have no *prima facie* indication that IBE as we have characterized it is a reliable argument form for truth-conducive inferences in science or in NM, even if the putative best explanation results in an increase in understanding.[14]

The prior claim assumes that one could understand a false explanation to greater or lesser degrees, or, in a stronger fashion, that a theory's being true is not a necessary or sufficient condition for understanding a theory. This is quite plausible. There are people who understand how Ptolemaic epicycles work better than others, despite that theory being false. Likewise with phlogiston theory, or any other now-debunked scientific theory. If all it took for an IBE to be successful was its engendering reports of increased understanding, then yes, there are plenty of successful IBEs and the standard of adequacy for the argument form is quite low. But I take it that is not all that scientists and naturalized metaphysicians expect of IBE. They don't just expect IBE to increase their understanding—they expect that theoretical virtues are confirmationally relevant, that the most explanatory theory is the most likely to be true and so the *best* theory to adopt among all competitors. But if understanding of a theory is separable from its being true or false, then there's a

[14] It may be that there are some forms of IBE which aren't predicated on an assessment of theoretical virtues. But I have trouble understanding how else we might appeal to the explanatory considerations brought to bear by a theory other than by appeal to the theoretical virtues.

conflation in what proponents of IBE take their arguments to be doing: increasing our understanding on the one hand and giving us a guide to the theory that is most likely to be true on the other.

The second problem with Ladyman's justification of IBE in the sciences is that it's not at all clear how inductively successful IBE in the sciences is, both in comparison to metaphysical IBE and on its own terms. This is because we don't have insight into the rate of success of IBE in science and metaphysics. This objection is analogous to the base rate fallacy objection found in the literature on the no-miracles argument and the pessimistic meta-induction.[15] The base rate fallacy is a common statistical fallacy where reasoners underweight or ignore the base rates of event probabilities and overweight individual information pertaining to the event in question. The classic example of the base rate fallacy comes from Kahneman & Tversky (1972):

> Two cab companies operate in a given city, the Blue and the Green (according to the color of cab they run). Eighty-five percent of the cabs in the city are Blue, and the remaining 15% are Green. A cab was involved in a hit-and-run accident at night. A witness later identified the cab as a Green cab. The court tested the witness' ability to distinguish between Blue and Green cabs under nighttime visibility conditions. It found that the witness was able to identify each color correctly about 80% of the time, but confused it with the other color about 20% of the time. What do you think are the chances that the errant cab was indeed Green, as the witness claimed?[16]

Many people, when faced with this problem, surmise that the chance of the cab's being green is identical to the witness's rate of correct color identification, that is, 80%. This is not the

[15] See Magnus and Callender (2004) for the argument as it pertains to the scientific realism debate.
[16] Kahneman and Tversky (1972).

case. The proper way to determine the likelihood that the cab was green given witness testimony of a green cab is given by Bayes' Rule:

$$Pr(A|B) = \frac{Pr(B|A) * Pr(A)}{Pr(B)}$$

Where $\Pr(A|B)$ is the conditional probability (i.e. the likelihood) of event *A* given event *B*, $\Pr(B|A)$ is the likelihood of event *B* given event *A*, and Pr(*A*) and Pr(*B*) are, respectively, the probability of *A* or *B*'s occurring independently. Filling in Bayes' Rule with the cab example we get:

$$Pr(G|g) = \frac{Pr(G|g)}{Pr(B|g)} = \frac{Pr(g|G)}{Pr(g|B)} * \frac{Pr(G)}{Pr(B)} = \frac{0.8}{0.2} * \frac{0.15}{0.85} = \frac{12}{17}$$

Where $(G|g)$ is the probability of the offending cab being green given a witness testimony of a green cab, $(B|g)$ is the probability of the offending cab being blue given a witness testimony of a green cab, $(g|G)$the probability of a witness testimony of a green cab given the cab's being green, $(g|B)$ the probability of a witness testimony of a green cab given the cab's being blue, and *Pr(G)* and *Pr(B)* being the base rate probabilities of a cab's being green or blue respectively. After applying Bayes' Rule as above we have:

$$Pr(G|g) = \frac{12}{12 + 17} = .41$$

This means that the probability of the offending cab's being green given a witness testimony of a green cab is not 80% but rather 41%. In one study by Bar-Hillel (1980) only about 10% of study participants gave an answer even roughly approximate to the correct calculation as given by Bayes' Rule.

The upshot of this discussion as it pertains to Ladyman is that we can't say that IBE has a track record of success in the sciences but not in metaphysics without knowledge of the base rate of successful IBEs found in both. Since we don't have the base rates available to us, we cannot assert that IBE in science has a track record of inductive success—perhaps the probability of an IBE's being correct in science is less than that of chance, which surely indicates a method that is not reliable. Furthermore, we are unlikely ever to discover the base rate of successful IBEs in science or metaphysics: There are too many IBEs in science to count and there are no clear criteria of verification for IBEs in metaphysics, so we can't calculate the base rates of successful IBEs in either discipline. While we can gesture toward greater levels of convergence in scientific consensus as a reason to suppose more success in inferring the best explanation and so a better track record in comparison to IBEs in metaphysics, this is ultimately an intuition that is in search of the probabilities that would justify it, since it could still be the case that IBE is unreliable (that is, not significantly better than chance) in both disciplines but to differing degrees.

One might respond that, so long as the explanatory power of a theory is conducive to its truth, what matters is that we have *some* set of weighting criteria, rather than any particular set of weighting criteria, if it is true that those who incorporate explanatory criteria into a Bayesian Inference scheme do better in general than those who strictly apply Bayesian Inference, as Lipton (2004), Douven (2013) and Tesic, Eva & Hartmann (2017) claim. Bayesian Inference uses Bayes' Rule to update the probability of a hypothesis as more evidence accumulates. Bayesian Inference takes the following form:

$$Pr(H|E) = \frac{Pr(E|H) * \Pr(H)}{\Pr(E)}$$

Where Pr $(H|E)$ is the probability of a hypothesis *H* given evidence *E*, Pr $(E|H)$ is the probability of evidence *E* given a hypothesis *H*, and Pr (*H*) and Pr (*E)* are, respectively, the probability of *H* and *E* being true independently.

The specifics of this proposal are technical, but, put briefly, the conclusion of the aforementioned articles is that versions of Bayesian inference that incorporate IBE into the calculation of prior probabilities perform better than standard Bayesian Inference. For example, Tesic, Eva & Hartmann (2017) claim that the IBE-supplemented version of Bayes' Rule strictly dominates the classical version in terms of picking the best hypothesis, that is, the following equation yields strictly positive values:

$$\Pr(H|E,X) - Pr(H|E)$$

Where $\Pr(H|E,X)$ is the probability of a hypothesis *H* given evidence *E* where hypothesis *H* is considered the best explanation *X* of evidence *E* and $Pr(H|E)$is the probability of a hypothesis *H* given evidence *E* where *H* is not considered the best explanation of *E*.[17] Douven (2013) makes a similar case through the use of various computer simulations, showing that an IBE-supplemented form of Bayesian inference does better than classic Bayesian inference in terms of choosing the theory most likely to be true, but also incurs a greater chance of choosing a false theory vis a vis classical Bayesian updating. On this account Bayesian IBE performs better on some specifications of an accuracy rule vis a vis the classical version of Bayes' Rule, which means at least in some cases it is an appropriate method to choose among theories, depending on one's tolerance for risk.[18]

These responses are wanting, however, because my objection holds even if explanatoriness really is truth-conducive and IBE respects and improves on strict application of Bayes' Rule[19] in the way that Lipton, Douven and Tesic, Eva & Hartmann claim that it does. My objection still holds because these accounts have still not shed any light on what the proper ordering and weighting of theoretical virtues is. This is explicitly acknowledged by Tesic, Eva, and Hartmann:

> It is important to note that we do not take ourselves to have contributed to the debate concerning the nature of the explanatory virtues. In particular, we have said nothing about what makes a hypothesis a good explanation of some given evidence.

---

[17] See Tesic, Eva & Hartmann (2017), Pgs. 18-20 for a proof of this claim.

[18] See Douven (2013) Pgs. 433-439 for more.

[19] IBE has to respect the rules of Bayesian conditionalization on posterior probabilities in order to avoid a dynamic Dutch book. See Teller (1973) and Van Fraassen (1989) for more.

> In the justification of our model, we equated explanatory virtue with the Bayesian information score because, *whatever one thinks about the nature of explanatory virtue*, it is natural to think that curves with lower information scores count as better explanations of the relevant evidence. We intend to remain ecumenical about the nature of explanatory virtue, and merely take ourselves to be providing conditions under which explanatory considerations can contribute to confirmation.[20]

Put briefly, the Bayesian information score is a criterion for model selection in Bayesian statistics. It is possible to increase the fit of a model by adding additional parameters to the model, but this increases the risk of overfitting. The BIS judges models on the basis of their likelihood functions, with a penalty term that increases as the number of parameters of the model increases. Those models with a lower BIS are taken to be superior to models with higher BIC, because they achieve similar or identical likelihood functions with fewer parameters and so are simpler, with less chance of overfitting to the data. But it must be said that even if we have an empirical criterion in the form of the BIS for adjudicating among models, this doesn't, by itself, entail that we have an uncontroversial method for adjudicating disputes about what is in fact the best explanation of the empirical data. This is because there are plenty examples where, other things being equal, the simpler explanation is not always the correct one, and the BIS has a marked favoritism for simplicity in parameters.[21] Nor is the BIS the sole criterion for model-selection—over ten different selection criteria are used with some regularity in statistical modeling. While it is among the most popular model selection criteria, its use doesn't by itself settle the question of how we could reliably determine which theoretical virtues should take precedence when informing our prior probabilities.

[20] Tesic, Marko; Eva, Benjamin & Hartmann, Stephan, Confirmation by Explanation: A Bayesian Justification of IBE. Pg. 13. Emphasis theirs.

[21] Presumably center of mass models of classical mechanics would have a lower BIS score than those that treat mass as distributed over an area (because there are fewer parameters in center of mass models, making them easier to calculate and functionally equivalent to distributed-mass models), but despite this center of mass models are *false*.

And this is important, because all Lipton, Douven, and Tesic et al. have shown is that there is *some* weighting of theoretical virtues that yields better outcomes (and, in Douven's case, only in certain specifications of an accuracy rule) than strict Bayesian conditionalization, not that all arrangements of the theoretical virtues yield better outcomes than strict Bayesian conditionalization. Explanatoriness being truth-conducive, then, is only a significant point in favor of IBE if we can specify the correct listing and hierarchy of theoretical virtues that we can then use to inform our prior probabilities. And currently we do not even have agreement over what candidate theoretical virtues are in fact theoretical virtues, and even less agreement on how they ought to be weighted.

These considerations point to a general problem with L&R's strategy: Any attempt to naturalize metaphysics by restricting theory choice preferences to just those that are naturalistically reputable is bound to fail as an appropriate criterion of theory choice so long as there is no antecedently specified objective ordering and weighting of theoretical virtues among working scientists. And the prospects of ever discovering an objective ordering and weighting of theoretical virtues among working scientists are dim. Scientists are, like everyone else, idiosyncratic in their hierarchizing and weighting of the theoretical virtues. Some prefer desert landscapes, others rainforests.

5 WHAT'S WRONG WITH HETEROGENEITY?

One may reasonably suspect that the claim that scientific IBE is unacceptably subjective applies with equal warrant to traditional Bayesian conditionalization. In fact, the problem of subjectivity might be worse in Bayesianism, insofar as scientific IBE places restrictions on the sorts of theories we might consider most likely (namely, those that satisfice a weighted plurality of explanatory virtues) whereas Bayesian conditionalization allows whatever assignment of prior probabilities one might desire; so long as one updates one's beliefs according to the dictates of Bayes' Rule, one is rational, irrespective of whether the assigned priors are wildly out of step with what is actually the case. This poses a problem, because if scientific IBE and Bayesianism both end up as unacceptably subjective, then we have no way to determine how we should update our choice of scientific theories in light of new evidence (because that is governed by Bayes' Rule) and, even more troubling, we have no guide to choosing a theory in general, since the two live methods are both considered unacceptably subjective. Note first that this analogy between IBE and Bayesianism ignores the existence of so-called "objective Bayesian" accounts like Rosenkrantz (1981), Berger (2006), and Weisberg (2009) that take this criticism of subjective Bayesianism as spot-on and so seek to identify objective constraints on the selection of priors. Whatever one thinks of this program, it is clear that some people take subjectivity in the method of theory-choice as a reason to doubt its effectiveness. What I want to claim, however, is that even the most subjective versions of Bayesianism that allow for the arbitrary assignment of values to priors are not objectionably subjective in the way that scientific IBE is.

How exactly is subjective Bayesianism less objectionably subjective than IBE? The answer lies in the idea that Bayesian priors (and likelihoods, *per* Earman [1992]) eventually "wash out" in the long run as new evidence accumulates. The mathematical result proving this claim is called the Bernstein-von Mises Theorem.[22] Put briefly, the theorem states that in the asymptotic limit Bayesian inference will approximate the posterior distribution (i.e. the probability distribution of an event or events conditional on newly observed data) whatever one's initial distribution of priors is, so long as the prior probability of a hypothesis *H* does not have an assignment of 1 or 0. For example, as the number of coin flips approaches infinity, Bayesian probability estimates converge on the relative frequencies of heads and tails, irrespective of one's initial priors for the likelihood of heads or tails, and so frequentist and Bayesian analyses of coin-flipping converge in the long-run.

[22] See https://www.encyclopediaofmath.org/index.php/Bernstein-von_Mises_theorem for a proof of the Bernstein-von Mises theorem and more besides.

It is important to note, however, that there are many situations in which the Bernstein-Von Mises Theorem does not hold (like in models with an infinite dimensional parameter, see Freedman [1999]), and so the Bernstein-Von Mises Theorem doesn't imply that the assigning of priors is irrelevant or arbitrary in every case.[23] What it does imply, however, is that there is a certain degree of convergence in subjective Bayesianism and that the arbitrariness for which subjective Bayesianism is often maligned for is rendered innocuous in some cases by the Bernstein-von Mises theorem. These considerations lead Douven to offer some criteria for the application of Bayesian inference:

> Bayesians, especially the more modest ones, might want to respond [to Bayesian IBE proponents] that the Bayesian procedure is to be followed if, and only if, either (a) priors and likelihoods can be determined with some precision and objectivity or (b) likelihoods can be determined with some precision and priors can be expected to "wash out" as more and more evidence accumulates or (c) priors and likelihoods can both be expected to wash out…In the remaining cases – they might say – we should simply refrain from applying Bayesian reasoning.[24]

Contra the objection, Bayesian reasoning is not an unacceptably subjective enterprise, insofar as we have a clear delineation of cases in which it should be applied (namely in situations where we can reasonably expect that both priors and likelihoods approximate relative frequencies or that they will wash out if they do not approximate relative frequencies)[25] and we have, in a non-trivial subset of cases, significant convergence upon results whatever our initial distribution of priors. This is admittedly a weak sense of convergence, given that convergence only happens asymptotically as the

[23] In any case where the Bernstein-von Mises Theorem doesn't hold, priors should not be expected to wash out, and so the selection criteria for priors significantly affects the outputs of a model unless they are considered "uninformative priors".
[24] Douven (2005). Pg. 342.
[25] Though see Gelman (2008) for an argument for the use of so-called "weakly informative priors" rather than priors that reflect posterior distributions or arbitrary priors that eventually wash out.

number of trials approaches infinity. But this convergence guarantees *some* consistency in results, under appropriate conditions, and so guarantees a degree of truth-conduciveness. IBE, on the other hand, currently has no prospects of convergence. As has been frequently mentioned, there is no widespread agreement over the proper listing of theoretical virtues in scientific or metaphysical practice, and there is even less agreement over how the theoretical virtues should be weighted in either discipline. And we don't have a particularly good reason to assume that our IBEs will converge in the long run until we have a specification of the base rates of success and the virtues on the basis of which we're judging.

It may be the case that if priors do in fact wash out then one is able to assign one's priors on the basis of explanatory considerations, and so one can freely use IBE as they see fit, since subjective Bayesianism places no restrictions on prior assignments at all. But this is *only* in the case where the priors do wash out in the (infinite) long-run, and in the short-term Bayesian IBE will achieve significantly different results than classical Bayesianism. Moreover, it is not guaranteed that the results will be better, given that the arguments brought to bear by Douven and Tesic et al. only prove that Bayesian IBE is superior (or at least preferable given a certain accuracy rule) to classical Bayesianism if we have the correct listing and organization of theoretical virtues. In cases where we can't expect priors to wash out, the use of IBE is inappropriate, because it doesn't give us any explicit and global constraints on the assignment of prior probabilities like objective Bayesian accounts do (just constraints on an idiosyncratic individual-by-individual basis).

6 WHITHER NATURALIZED METAPHYSICS?

If L&R's account of NM fails, are we left with no principled distinction between naturalized and non-naturalized metaphysics? I think that conclusion is too hasty. What the foregoing considerations requires upon the part of the naturalized metaphysician is a reorientation away from a focus on a defining a naturalist criterion of theory choice and towards a view that sees metaphysical practice as fundamentally about constructing physically possible models of reality. This view is a descendent of various "metaphysics-as-modeling" accounts (I will refer to these views as MAM, and to my "naturalized metaphysics-as-modeling" as NMAM) recently advanced by Peter Godfrey-Smith (2006), Laurie Paul (2012), Timothy Williamson (2017), and others. NMAM, in contrast to MAM, comes with greater constraints on the production of models, requiring them to not be metaphysically possible but also *physically* possible, which justifies appending the word "naturalist" to the practice of metaphysical model-building. This view can salvage one of the aims of *Every Thing Must Go*, namely, supplying an account of metaphysical practice as strongly constrained by the practice of science. I will first define some key terms and then explore how the similarities and differences between my view of naturalized metaphysics and L&R's view.

First, a definition of what it is to be a model is in order. My definition of model is very much along the lines of Peter Godfrey-Smith's:

> A rough definition of the relevant sense of "model" can be given as follows: A model is an imagined or hypothetical structure that we describe and investigate in the hope of using it to understand some more complex, real-world "target" system or domain. Understanding is achieved via a resemblance relation, that is, some relevant similarity, between the model and the real-world target system… Model-based science takes an *indirect* approach to representing complex or unknown processes in the real world. The modeler's first move is to specify and try to understand a hypothetical structure, often using mathematical methods. It is a separate question—and often a very subtle one—to work out what sort of similarity there is between the model and events and processes in the real world.[26]

[26] Godfrey-Smith (2006). Pg. 7.

Scientific models, then, are structures that are designed to simplify the workings of an actually existing system by isolating some set of relevant variables, whose outputs are then used to inform further theorizing about the target system.[27] Metaphysical models are largely similar, but differ in that they are speculative, non-testable *extensions* of scientific theory—ways the world could be, given what we know about physics.[28]

Now I will define metaphysical possibility and physical possibility. Unless one is a dyed-in-the-wool constructive empiricist, it is eminently plausible that the world could have been a different way than it currently is. It is not a mystery or a surprise that the table on which I am writing could have been two feet to the left, but in fact isn't. And if the world could be different in these small ways, it seems, by parity of reasoning, that it could be different in significant ways—that, say, the value of the fine structure constant could be different, or that the speed at which objects in free-fall descend due to gravity could be different, and so on. For a state of affairs to be metaphysically possible, then, is for it to be a possible world in the classic sense of Kripke (1980) and Lewis (1986). It is (perhaps) metaphysically possible that there exists a world with only two homogenous spinning disks in it[29]; it is also metaphysically possible that there exists a world where the desk I am currently writing on is two feet to the left. It is not metaphysically possible, so far as I can tell, that there exists a world with round squares. Without trying to take sides on the debate over the ontology of possible worlds, I take metaphysical possibility to be the following:

[27] See Giere (2002) for an overview of how modeling is applied in the sciences, and what this means for the philosophy of science generally.
[28] Testability is what separates metaphysical models from highly abstract scientific models like Loop Quantum Gravity. LQG makes predictions, and so is testable, but all of its predictions are made at the Planck scale, and our particle colliders are very far away from achieving such high-energy states.
[29] See Armstrong (1980) for the advent of this much discussed thought experiment and its relation to Lewis' Humean Supervenience.

> Metaphysical Possibility: A state of affairs *a* is metaphysically possible if and only if at least one possible world *w* contains *a*.

The distinction between metaphysical and physical possibility concerns what science tells us about the structure of our actual world. While it is metaphysically possible that there could be a world with two homogenous spinning discs, it is not physically possible, given the laws of physics as we currently understand them. *Our* world could not have contained only two homogenous spinning discs, unless the laws of physics were significantly different from what they actually are. It is physically possible, however, that the table on which I am writing could have been two feet to the left; it is also physically possible that I could phase through my (solid and intact) chair due to the effects of quantum tunneling, though the chances of this happening are so vanishingly small that they are effectively zero. On the other hand, it is not physically possible, given our current understanding of General Relativity, for information to be transmitted faster than the speed of light, and so faster-than-light communication is not physically possible, though if the laws of physics were somewhat different, it perhaps would be possible.[30] So while it may be metaphysically possible that there exists a world with two homogeneous spinning disks or that there exists a world with faster-than-light signal transmission, neither appears physically possible given what we know about physics. The definition of physical possibility then, is:

> Physical Possibility: Some state of affairs *a* is physically possible if and only if some plausible model of actual physics includes *a*.[31]

---

[30] Though some processes, like the phase velocity of a wave or the group velocity of a wave, can propagate at a speed greater than the speed of light in a vacuum, they do not transmit any information. General Relativity doesn't claim that processes cannot propagate faster than *c*, only that signal transmission will always be less than or equal to *c*.

[31] This distinction between metaphysical and physical possibility is expanded on from some comments made by Tim Maudlin. See Maudlin (2007), pg. 187, for more.

What is a "plausible model of actual physics"? A plausible model of actual physics is a physical model that is isomorphic to some set or subset of physical phenomena. String Theory and Loop Quantum Gravity are plausible models of actual physics, insofar as they are self-consistent models of physical phenomena that subsume and unify QM and GR. Applied mathematical theory (e.g. continuum mechanics) also yields plausible models of actual physics, though at a more restricted scale and scope than that of GR or QM. Since it is presumably the case that any chemical, biological, or social-scientific model won't posit entities or laws that contradict physics (unless they incorporate deliberate fictionalizations), they are also physically possible, and are under the purview of naturalized MAM.

How exactly does NMAM differ from L&R's view of NM? It differs in three significant respects: First, NMAM does not motivate any of its naturalism by way of appeal to the preferences of working scientists. Rather, NMAM is motivated by the recognition that those models that are of use to the practice of science are those of the actual world, not any nearby or more distant possible worlds. Whatever the merits of theorizing about a world with two homogeneous spinning discs, it doesn't appear that this kind of theorizing offers any kind of heuristic or methodological value to science as it is currently practiced. Since naturalized metaphysics takes being of heuristic and methodological value to scientific inquiry as its highest desideratum, those aspects of metaphysics that are taken to not be useful for the purposes of scientific inquiry should be discarded. On this account, the main difference between Loop Quantum Gravity and a world with two homogenous spinning disks is that the former is a physically possible model of reality and is likely (but not guaranteed) to be in-principle empirically verifiable, while the latter is not a physically possible model of reality and is likely (but not guaranteed) to be in-principle empirically unverifiable. It is at least plausible that those models that are likely in-principle empirically unverifiable offer little to no methodological or heuristic value to scientific inquiry, insofar as scientists are loath to adopt research programs that cannot ever be tested.

Second, NMAM, unlike L&R's view of naturalized metaphysics, does not endorse outright belief in the models that it generates. Because models are not in the business of direct theorizing—models are not meant to be exhaustive explanations and descriptions of actual phenomena, only partial representations of a system for the purpose of further study—they also are (when properly interpreted) not meant to engender belief in the model *per se*. Models are pragmatic devices, and are free to simplify, idealize, and fictionalize, so long as they retain sufficiently similarity to the target system. [32]

Finally, NMAM also offers a strong view of science's constraint on metaphysical practice while avoiding the untenable appeal to a naturalist criterion of theory choice. For NMAM, what constrains metaphysical practice is the requirement that metaphysical models be physically possible, not just metaphysically possible. This is a fairly stringent criterion insofar as contemporary physics enforces many constraints on what counts as physically possible. The motivation for this constraint is, again, that science cares about the actual world, not possible worlds.[33] On this view, L&R's metaphysical view—what they call "information-theoretic structural realism"—is (by my lights) a physically possible model, and so merits consideration and analysis to see how well it fits with what we know to be the case and what fruits it might bear. But it is not any better, in an absolute sense, than any other model that satisfies the constraint of physical possibility. We might prefer one model

[32] Of course, there is great controversy over what "sufficient similarity to the target system" entails, especially in cases of idealization and fictionalization. Those concerns are beyond the ken of this paper, however.

[33] And also because we have much more reliable knowledge about the actual world versus those that are merely possible, but not actual. But this is a separate argument, and though I am sympathetic to L&R's critique of analytic metaphysics, to explain and justify why this is the case would take us too far beyond the subject of this paper.

to another for a whole host of reasons, and some of those reasons might include judgments about how accurately they actually do model reality. But, assuming that we can't determine which model of reality (say, Loop Quantum Gravity vs. String Theory) is the "best" on the basis of its theoretical virtues alone (and the preceding sections should indicate that a naturalized IBE probably can't do this), then naturalized MAM encourages a pluralism regarding model selection and model-building. This is a feature, not a bug. Models are by necessity simplified and incomplete, and multiple models can represent different views of the same phenomenon and convey information regarding those different aspects. A healthy pluralism, then, is methodologically beneficial, and, because of the constraint of physical possibility, we can rest assured that our metaphysical inquiry won't venture too far into the thicket of possible worlds and thereby lose its usefulness to scientific inquiry.

There are two major differences between modeling in NMAM and modeling in the sciences that need to be addressed, however. The first difference between scientific models and NMAM models is the former need not always satisfy the criterion of physical possibility that I've forwarded as a criterion of model building in NMAM. Simple models of a pendulum accurately predict the motion of pendulums in our world, but do not contain a coefficient for friction, meaning that simple model pendulums are perpetual motion machines, which are physically impossible. This is not a problem for NMAM, if one acknowledges that the purposes of modeling in science and NMAM are somewhat different. The purpose of modeling in science is, again, to be able to gather more information about a target system by simplifying, idealizing, or fictionalizing its workings through the use of an indirect structure that captures some of the relations among the target system's parts. The point of modeling in NMAM is to encourage, *inter alia*, the possible conceptual and empirical extensions of already existing physical theory, and by doing so encourage scientists and philosophers to pursue new avenues of inquiry. NMAM isn't about information-gathering so much as it is about encouraging the furtherance of existing research programs[34] and the generation of new ones by working scientists. The second difference is that models in science are usually *local* rather than *global*. This characteristic of models is foregone in NMAM, because NMAM's purpose is heuristically useful speculation about the way our actual world could possibly be. If NMAM were meant to generate true facts about the world, rather than serve as a methodological handmaiden to scientific inquiry, then the constraint of locality would be more well-motivated, as it is in scientific modeling.

So far, I have differentiated NMAM from L&R's view of NM and from standard MAM, but I haven't given an example that I take to be representative of what NMAM is. I will now briefly present an exemplar of NMAM, Newtonian atomism, and hopefully illustrate through this example how NMAM can aid scientific inquiry.

Newton's atomism is situated within a lively tradition of mechanistic atomist philosophy that spans from the 17th to 19th centuries, a tradition that was dominant until the advent of Special Relativity in the 20th century. Classical mechanical philosophy, advocated by Robert Boyle, Thomas Hobbes, Rene Descartes, Pierre Gassendi, and others, treats all of nature as consisting in the movements of homogeneous matter, whose movements are defined by appeal to micro-scale properties found at ever-decreasing scales of extension. While not all mechanical philosophers were atomists (Descartes believed that matter consisted in extension and was infinitely divisible, for example), many natural philosophers and scientists of the era found atomism and mechanical philosophy kindred intellectual spirits and endorsed the former as a model of the latter. The most notable advocate of this view before Newton was Robert Boyle, who endorsed a variant of atomism known as corpusclarianism. But Newton's atomism was different from his predecessors, insofar as it was explicitly informed by his foundational work on dynamics, forces, and the laws of motion,

[34] I mean the term "research program" roughly in the sense of Lakatos (1978).

whereas previous mechanistic models like Boyle's had trouble explaining all of natural phenomena through only the size, shape, and motion of individual particles without the assistance of Newtonian Dynamics. Newton justified his atomism by appealing to atomism's ease in incorporating phenomena into the mechanistic framework.

> All Bodies seem to be composed of hard Particles: For otherwise Fluids would not congeal; as Water, Oils, Vinegar, and Spirit or Oil of Vitriol do by freezing; Mercury, by dissolving the Mercury and evaporating the Flegm…Even the Rays of Light seem to be hard Bodies; for otherwise they would not retain different Properties in their different Sides. And therefore Hardness may be reckon'd the Property of all uncompounded Matter. At least, this seems to be as evident as the universal Impenetrability of Matter…Particles being Solids, are incomparably harder than any porous Bodies compounded of them; even so very hard, as never to wear or break in pieces…While the Particles continue entire, they may compose Bodies of one and the same Nature and Texture in all Ages: But should they wear away. Or break in pieces, the Nature of Things depending on them, would be changed…the Changes of corporeal Things are to be placed only in the various Separations and new Associations and Motions of these permanent Particles; Compound Bodies being apt to break…These principles I consider, not as occult Qualities, supposed to result from the specific Forms of Things but as general Laws of Nature, by which the Things themselves are form'd; their Truth appearing to us by Phenomena, though their causes be not yet discover'd.[35]

Note that Newton's justification of atomism makes no reference to any experiments. Rather, Newton appealed to the fact that 1. Atomism was empirically adequate, accommodating the

[35] *Opticks*, *Query 31*, pgs. 389, 400-401.

scientific phenomena to which it was taken to apply (e.g. the freezing of fluids, evaporation of solids in liquids); and 2. That it cohered with a mechanistic account of nature.

It is important to stress again that, at that time, there had been no experimental confirmation of the atomist thesis. In fact, it would take until Roger Dalton formulated his version of atomism before atomism became a live empirical hypothesis that made testable predictions, and atomism didn't really assume the mantle of a *bona fide* scientific research program until Maxwell presented his kinetic theory of gases, which eventually culminated in a verification of atomism via Jean Perrin's experiments with Brownian motion in the early 20th century.[36] Because of these considerations, I take Newtonian atomism to be an exemplar of NMAM for its time period. It constructed a physically possible and general model of reality, in advance of empirical confirmation or disconfirmation, that served as a heuristically useful model for future chemists and physicists despite its lack of empirical confirmation. Because of its heuristic value, it significantly aided the burgeoning empirical research program of atomic theory, which is now no longer just a physically possible model of reality but also well-confirmed science consensus. The cash value of NMAM, then, lies in the construction of physically possible models of reality that are of heuristic and/or methodological benefit to existing or nascent scientific research programs.

7 CONCLUSIONS

After profiling Ladyman & Ross' account of naturalized metaphysics, I argued that their justification for their main methodological prescription, the PNC, was ill-motivated, predicated as it was on an ill-conceived deference to the theory-choice preferences of working scientists. I then explored how L&R's deference to the theory-choice preferences of working scientists rendered the prospect of naturalized metaphysics incoherent, because taking L&R's principle of deference

[36] See Chalmers (2009, 2014) for more on the atomist thesis from the 17th to 20th centuries.

seriously forces us to let in all of the theory choice preferences that scientists do in fact have, and those are not significantly different than the ones the metaphysician has. I subsequently critiqued Ladyman's commitment to treating IBE as truth-conducive in scientific inquiry and claimed that any view of naturalized metaphysics that attempted to offer a naturalized criterion of theory choice was bound to fail. I then answered an objection that analogized IBE and Bayesian Conditionalization and claimed that the former but not the latter was unacceptably subjective because IBE so far has had no reasonable progress toward convergence. Finally, I offered an alternate view of naturalized metaphysical practice that sees the purpose of metaphysics as one of constructing physically possible models of reality, a view that I take to salvage many of Ladyman & Ross' methodological prescriptions without relying on a view of scientific constraint that essentially refers to the practices of working scientists. I closed by presenting Newtonian atomism as a paradigm case of NMAM.

REFERENCES

Armstrong, David (1980). Identity Through Time. In Peter van Inwagen (ed.), *Time and Cause: Essays Presented to Richard Taylor*. Reidel. pp. 67-78.

Bar-Hillel, Maya (1980). The base-rate fallacy in probability judgments. *Acta Psychologica* 44 (3):211-233.

Berger, James. The case for objective Bayesian analysis. Bayesian Anal. 1 (2006), no. 3, 385--402. doi:10.1214/06-BA115. https://projecteuclid.org/euclid.ba/1340371035

Bryant, Amanda (2017). Keep the Chickens Cooped: The Epistemic Inadequacy of Free Range Metaphysics. *Synthese*:1-21.

Chalmers A. (2009) Newton's Atomism and its Fate. In: The Scientist's Atom and the Philosopher's Stone. Boston Studies in the Philosophy of Science, vol 279. Springer, Dordrecht

Chalmers, Alan, "Atomism from the 17th to the 20th Century", *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/atomism-modern/>.

Douven, Igor (2002). Testing Inference to The Best Explanation. *Synthese* 130 (3):355-377.

(2013). Inference to the Best Explanation, Dutch Books, and Inaccuracy Minimisation. *Philosophical Quarterly* 63 (252):428-444.

Douven, Igor. (2005). Wouldn't it be lovely: Explanation and scientific realism. *Metascience* 14:338-343.

Earman, John (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. MIT Press.

Freedman, David. Wald Lecture: On the Bernstein-von Mises theorem with infinite-dimensional parameters. Ann. Statist. 27 (1999), no. 4, 1119--1141. doi:10.1214/aos/1017938917. https://projecteuclid.org/euclid.aos/1017938917

Gelman, Andrew; Jakulin, Aleks; Pittau, Maria Grazia; Su, Yu-Sung. A weakly informative default prior distribution for logistic and other regression models. Ann. Appl. Stat. 2 (2008), no. 4, 1360--1383. doi:10.1214/08-AOAS191. https://projecteuclid.org/euclid.aoas/1231424214

Giere, Ronald N. (2002). How models are used to represent reality. *Philosophy of Science* 71 (5):742-752.

Godfrey-Smith, Peter (2006). Theories and Models in Metaphysics. *The Harvard Review of Philosophy* 14 (1):4-19.

Keas, Michael N. (2017). Systematizing the theoretical virtues. *Synthese*:1-33.

Kripke, Saul (1980). *Naming and Necessity*. Harvard University Press.

Ladyman, James (2012). Science, metaphysics and method. *Philosophical Studies* 160 (1):31-51.

Ladyman, James & Ross, Don (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford University Press.

Lakatos, Imre (1978). The Methodology of Scientific Research Programmes. Cambridge University Press.

Lewis, David K. (1973). *Counterfactuals*. Blackwell.

Lewis, David K. (1986). *On the Plurality of Worlds*. Wiley-Blackwell.

Lipton, Peter (2004). *Inference to the Best Explanation*. Routledge/Taylor and Francis Group.

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

Magnus, P. D. & Callender, Craig (2004). Realist Ennui and the Base Rate Fallacy. *Philosophy of Science* 71 (3):320-338.

Maudlin, Tim (2007). *The Metaphysics Within Physics*. Oxford University Press.

Newton, Isaac (1704). *Opticks*. Dover Press.

Paul, L. A. (2012). Metaphysics as modeling: the handmaiden's tale. *Philosophical Studies* 160 (1):1-29.

Peirce, C.S. Project, Peirce Edition (ed.) (1998). *The Essential Peirce, Volume 2: Selected Philosophical Writings (1893-1913)*. Indiana University Press.

Quine, W. V. (1966). *The Ways of Paradox, and Other Essays*. Harvard University Press.

Rosenkrantz, Roger D. (1981). Inference, Method and Decision: Towards a Bayesian Philosophy of Science. *Journal of Philosophy* 78 (6):356-367.

Teller, Paul (1973). Conditionalization and observation. *Synthese* 26 (2):218-258.

Tesic, Marko ; Eva, Benjamin & Hartmann, Stephan, Confirmation by Explanation: A Bayesian Justification of IBE.

Van Fraassen, Bas C. (1989). *Laws and Symmetry*. Oxford University Press.

Weisberg, Jonathan (2009). Locating IBE in the Bayesian Framework. *Synthese* 167 (1):125-143.

Williamson, T. (2017). Model-Building in Philosophy. In Philosophy's Future (eds R. Blackford and D. Broderick). doi:10.1002/9781119210115.ch12