Article:
Bueno, O and Shalkowski, SA (2020) Troubles with Theoretical Virtues: Resisting Theoretical Utility Arguments in Metaphysics. Philosophy and Phenomenological Research, 101 (2). pp. 456-469. ISSN 0031-8205

https://doi.org/10.1111/phpr.12597

TROUBLES WITH THEORETICAL VIRTUES: RESISTING THEORETICAL UTILITY ARGUMENTS IN METAPHYSICS*

Otávio Bueno

University of Miami

Scott A. Shalkowski

University of Leeds

Abstract

In this paper we examine theoretical utility arguments in metaphysics. While philosophers claim a procedural continuity with science when using such arguments, we argue that examining famous instances from the history of science expose their fundamental flaws. We find that arguments from theoretical utility invoke considerations that are not truth conducive‘ and that justifications for claims that a theory possesses theoretical virtues often assume the truth of the theory such virtues are supposed to support. We conclude that theoretical utility arguments provide no epistemic grounds for metaphysical inquiry.

Keywords: theoretical utility, consistency, coherence, explanatory power, indispensability, inference to the best explanation, truth.

1 Introduction

The metaphysicians' project is to delineate what there is and how it is, but it must be distinguished sharply from the scientists' project which, at least for naturalists, is fairly characterized in precisely the same terms. Metaphysicians and at least some scientists seek informative, illuminating explanations and both seek some grasp of what is more and what is less fundamental. Scientists have to hand well-known empirical methods for separating the scientific wheat from chaff, even though all recognize these to be fallible methods that underdetermine defeasible conclusions.

Lacking the luxury of empirical methods, metaphysicians nevertheless look for accountings of what there is and how it is that are "deeper" than those available from empirical methods. Metaphysicians must provide grounds for thinking that their chosen methods or patterns of inference have some tendency to direct us toward the truth—such patterns must be truth conducive. The crucial inferences conclude that

*We are grateful to Steven French for particularly helpful discussions on the issues examined in this work.

theories are (more likely to be) true (compared to competitors) when the theories themselves have certain internal characteristics, i.e., when they have some sufficiently high degree of theoretical virtue, such as being simple, unified, expressive, and explanatory. We will argue that not only would this mode of inference have led us astray had it been used in some famous empirical contexts, there are no non-question-begging reasons for thinking that the most commonly cited theoretical virtues are truth conducive at all. We conclude that this mode of justification is a poor method for well-grounded metaphysics.

2 Metaphysical Theorizing and Theoretical Virtues

Scientists are often presented as positing new ontologies (from atoms to black holes) when the theories making those posits have a number of theoretical virtues. Some metaphysicians then seem to assume that if something is good for the sciences, it should be good for metaphysics as well, *ceteris paribus*. Sufficient epistemic merit for arguments from theoretical virtues in metaphysics is claimed on the basis of the sufficiency of similar arguments in science and mathematics.

The development of both scientific and philosophical theories is a matter of amassing data and then constructing theories that accommodate the data. Speaking for many in the metaphysics community, Ted Sider, John Hawthorne, and Dean Zimmerman write:

> Scientists must regularly choose between many theories that are consistent with the observed data. Their choices are governed by criteria like simplicity, comprehensiveness, and elegance. This is especially true in very theoretical parts of science, for instance theoretical physics [...].
>
> [...] Just like scientists, metaphysicians begin with observations, albeit quite mundane ones: there are objects, these objects have properties, they last over time, and so on. And just like scientists, metaphysicians go on to construct general theories based on these observations, even though the observations do not logically settle which theory is correct. In doing so, metaphysicians use standards for choosing theories that are like the standards used by scientists (simplicity, comprehensiveness, elegance, and so on) (Sider et al., 2008, p. 6).

The emphasis on the continuity between science and metaphysics shapes more generally Sider's own epistemology of metaphysics. He correctly highlights the inherently fallible nature of metaphysical inquiry, and identifies the continuity between the two domains because both use theoretical virtues in theory choice.

> The epistemology of metaphysics is far from clear; this any metaphysician should concede. For what it's worth, as a general epistemology of metaphysics I prefer the vague, vaguely Quinean, thought that metaphysics is continuous with science. We employ many of the same criteria—whatever those are—for theory choice within metaphysics that we employ outside of metaphysics. Admittedly, those criteria give less clear guidance in metaphysics than elsewhere; but there's no harm in following this argument where it leads: metaphysical inquiry is by its nature comparatively speculative and uncertain (Sider, 2011, p. 12).

Sider then generalizes the point to support his own preferred ontology: structure. One's best theory exhibits the relevant theoretical virtues (Sider singles out simplicity, explanatory power, and unification). He then follows Quine's advice: the ontology of one's best theory should be one's ontology. He maintains that it is not only the ontology that is read from one's best theory; the same goes for the theory's ideology. The theoretical resources articulated by one's best theory are likely to capture the relevant structure in the world; they "carve at the joints". Sider writes:

> This Quinean thought suggests an epistemology for structure in particular. Quine's advice for forming ontological beliefs is familiar: believe the ontology of your best theory. Theories are good insofar as they are simple, explanatorily powerful, integrate with other good theories, and so on. We should believe generally what good theories say; so if a good theory makes an ontological claim, we should believe it. The ontological claim took part in a theoretical success, and therefore inherits a borrowed luster; it merits our belief. This all is familiar; but a believer in structure can say more. A good theory isn't merely likely to be true. Its ideology is also likely to carve at the joints. For the conceptual decisions made in adopting that theory—and not just the theory's ontology—were vindicated; those conceptual decisions also took part in a theoretical success, and also inherit a borrowed luster. So we can add to the Quinean advice: regard the ideology of your best theory as carving at the joints. We have defeasible reason to believe that the conceptual decisions of successful theories correspond to something real: reality's structure (Sider, 2011, p. 12).

Timothy Williamson similarly emphasizes the role that theoretical virtues play in identifying what a good theory is, while also connecting the goodness of an explanation to its truth. As he notes:

> [...] the more T has the intrinsic virtues of a good theory, the better (*ceteris paribus*). It should be elegant and unified, not arbitrary, gerrymandered, *ad hoc*, or messily complicated. It should be informative and general. In brief, it should combine simplicity with strength (Williamson, 2016, p. 266).

Tellingly, Williamson presents theoretical virtues as being intrinsic to the theory in question. This makes theory choice by theoretical virtues tractable: one need not consider factors outside the theory to determine whether it is virtuous or not. Both properly and troublingly, Williamson frames the explanatory virtue in terms of *potential* explanations, which are those that '*would* explain the evidence *if [they] were true*' (2016, p. 266). Properly, because it correctly highlights that the forms of argument become more tractable if they emphasize the potential nature of explanation, but troubling, because it makes salient the first important oddity of those forms: the inference that a theory *is* the correct explanation because it *would be* the best explanation.

Proponents of arguments from theoretical virtues face a dilemma. Either theoretical virtues are intrinsic or they are not. If they are intrinsic, since truth is not an intrinsic feature of a theory, there is no guarantee that the satisfaction of the theoretical virtues will make the theory more likely to be true. Below, we cite instances where the world did not cooperate with virtuous theories. If theoretical virtues are instead extrinsic, then determining that a theory is virtuous requires accounting for what goes on beyond the theory and in the world, thus losing the entire point of appealing to the virtues

in the first place, which is breaking deadlocks that remain from considering extrinsic factors, as in empirically equivalent theories.

Juha Saatsi has identified, under the heading of "explanationism", a cluster of approaches that highlight the role of inferences from theoretical virtues that focus primarily on how well a theory would explain the relevant phenomena (Saatsi, 2017). Such approaches then use the best explanation as a basis for believing the relevant theory. These inferences are common across a wide range of theories in metaphysics. They are used to address topics as diverse as the existence of laws of nature (Armstrong, 1983, 1997; Bigelow and Pargetter, 1990; Dorato, 2012), the existence of mathematical objects (Colyvan, 2006; Bigelow and Pargetter, 1990), and of abstract objects more generally (Swoyer, 2008), as well as defences of scientific realism (Psillos, 2005; Ellis, 2009), and the articulation of views in mereology (Bigelow, 2010), ontology of mind (McLaughlin, 2010), and in the epistemology of modality (Biggs, 2011; Fischer, 2017), all the way to the very possibility of ontology itself (Swoyer, 1999), and the methodology and epistemology of metaphysics (Paul, 2012).

Some may urge ranking theories on the basis of their respective theoretical virtues as the grounds for theory choice only when theories are empirically equivalent. Empirical considerations typically trump theoretical constraints, thus preempting any need for the use of theoretical virtues. Thus, only given two or more empirically equivalent theories would one select the theory that satisfies the theoretical virtues to the greatest degree. Since typical metaphysical theories make no empirical difference, they are all empirically equivalent, and choosing theories on the basis of theoretical virtues is appropriate across the board.

Interestingly, empirical equivalence figures explicitly only in the passage by Sider, Hawthorne, and Zimmerman. Perhaps the others simply assume the empirical equivalence of metaphysical theories.

Why, though, should one accept this constraint on the relevance and use of theoretical virtues in the first place? It is doubtful that all metaphysicians accept the constraint that theoretical virtues be evaluated only when theories are empirically equivalent. Some metaphysical theories are not even empirically adequate and hence cannot be empirically equivalent to theories that are compatible with the empirical information. For instance, certain metaphysical theories conflict with well-established results from quantum mechanics or relativity theory. Some require that identity be applicable to every object and, thus, are inconsistent with certain formulations of non-relativistic quantum mechanics. Others assume that in principle any material object could accelerate to speeds beyond that of light, thus conflicting with relativity theory.

Moreover, if theoretical virtues are indeed truth conducive—something we will argue against below—why should empirical considerations always trump them anyway? If a simple, unified, and explanatory theory posits a certain entity, why should one revise the theory, if such entity is not found? Why not challenge the standing of the experiments in which the entity in question has not been detected? Or, why not criticize the interpretations of such experiments so that the established results can be reconciled with the virtuous theory? If empirical considerations are always given the upper hand, what grounds are there for thinking that the virtues are indicators of truth at all? Exactly why can they never aggregate to outweigh the significance of at least some empirical data? At the very least, maintaining the truth conduciveness of theoretical virtues requires an account of exactly how they have such epistemic merit and where they fit the hierarchy of truth-conducive considerations.

In what follows, we do not assume that theoretical virtues should be applied only to empirically equivalent theories. In fact, we argue that in a number of prominent cases in the history of science, had theoretical virtues been used as a guide to theory choice, scientists would have moved away from the truth. This casts doubt on the adequacy of the entire theoretical virtues approach. We then argue that theoretical virtues cannot play the role of being a guide to truth in metaphysical theorizing either. Throughout the discussion, the central difficulty is that such virtues are not truth conducive.

3 The Basic Form

The largely assumed and unstated form of arguments from theoretical virtues is some variation on *modus ponens*:

(1) If a theory T is more virtuous (i.e., it satisfies more theoretical virtues) than any of its relevant competitors, then T is true.

(2) T is more virtuous than any of its relevant competitors.

(3) T is true.

Of course, real-life instances of the argument are slightly more subtle. (1) is usually qualified. T is only more likely to be true than its less virtuous competitors. The 'more likely' is likewise qualified to mean that there is more reason to believe T rather than any of its competitors, since, in many instances, it is hard to determine the objective probabilities for theories. Thus, it is difficult to use effectively such probabilities in rational theory choice.

This form of argument has been widely adopted in metaphysics. Consider, for instance, the use that Williamson makes of it. After claiming that the more a theory T 'has the intrinsic virtues of a good theory, the better (*ceteris paribus*)', he remarks:

> When a theory T scores highly enough as a potential explanation of our evidence E, and better than its rivals [that is, when T is virtuous], we may infer T from E by inference to the best explanation (Williamson, 2016, p. 266).

Clearly, Williamson is endorsing here the first premise of the theoretical virtues argument (albeit he is casting his remark as part of an abductive inference form, to which we return later, rather than a deductive one).

These qualifications might induce one to think that the basic form above is too basic. Perhaps we should build into the form itself issues of reasons for belief.

(1′) If T is more virtuous than any of its relevant competitors, then there are good reasons to believe T.

(2′) T is more virtuous than any of its relevant competitors.

(3′) There are good reasons to believe T.

What we gain on the swings, we lose on the roundabouts. (1′) presents us with a bit of misdirection, since, in contrast to (1), it no longer focuses on the truth of the relevant theories. Much of so-called analytic metaphysics styles itself as engaged in truth-directed theorizing. The project is to discern how things are and reasons to believe how *these* things are is to be informed by how *those* things are. Talk of theories and their truth or falsity is just talk of how things are in linguistic guise. Good reason

for belief absent the implication that what is believed is properly taken to be the case typically is not the goal of those arguing from theoretical virtues.

It is no violence, then, to the use to which inferences from theoretical virtues are put to treat the original *modus ponens* argument as sufficiently correct. No part of the following discussion exploits complaints about T being the best of a bad lot or that T may still be quite unlikely, given what is to be said for its competitors and the sum of their "likelihoods". We will examine critically the assumption that scientific practice vindicates such reasoning in empirical contexts. If it is not vindicated there, then the claim that at least this mode of philosophical inference is continuous with the sciences falls at the first hurdle. We will examine in turn the merits of each premise, arguing that (1) is false and that no dialectically-respectable grounds are available for (2).

4 The Conditional Premise

We question both the extent to which scientific practice itself relies on such theoretical virtue arguments (beyond providing pragmatic reasons for acceptance of the relevant theories) and whether these virtues legitimately provide grounds for commitment to the existence of the ontology that is thereby postulated. First, scientific theory choice typically appeals to theoretical virtues in comparative contexts when theories are roughly empirically equivalent. In this case, theories are chosen primarily for pragmatic reasons; empirically there is no differentiating them. But pragmatic reasons are not epistemic: they provide reasons for preferring to work with a theory rather than reasons for how well the theory represents the world. Pragmatic choices like these provide no guidance to metaphysicians who intend at least to approximate the truth in matters metaphysical. (For an account of theoretical virtues as pragmatic rather than epistemic—that is, as concerned with the users of the theory, their preferences and predilections, rather than with the relation between theory and the world—see (van Fraassen, 1980, pp. 12–13); for a discussion of this point in the context of a theory of possible worlds, see (Bueno and Shalkowski, 2015).)

As argued in §2, metaphysicians maintaining that theoretical virtues have epistemic significance owe us some argument showing how these matters manage to track reality. On pain of circularity, the metametaphysical argument cannot be that it is theoretically more virtuous to maintain that theoretically more virtuous metaphysical theories are more likely to be true. A direct argument is needed for the conclusion that simplicity, unification, and expressive and explanatory power (or any other theoretical virtues to which one might like to appeal) are indeed truth conducive. It will, furthermore, not do to argue that theoretically more virtuous scientific theories have shown themselves to be closer to the truth than their less virtuous competitors, for two reasons. The first and most philosophically-general reason is that as with all generalization strategies, the generalizing must be warranted. Since empirical matters are so narrow when compared to the panoply of metaphysical possibilities, there is not yet good reason to think that the narrow range of possibilities that are currently consistent with our best physical theories are sufficiently like the rest of the range of possibilities to warrant generalizing empirical methods for metaphysical purposes. This reason speaks to the alleged difference between scientific and metaphysical projects. The second reason speaks to an error in the key part of the relevant metaphysical myth, namely that science's drive toward truth is guided by comparative virtues.

4.1 Considerations from the Sciences

Consider the celebrated case of the transition from Ptolemy's to Copernicus's theory in astronomy, of which Kuhn (1957) provides a rich examination and Cohen (1983) offers a useful discussion. The standard ways of assessing the relative virtuousness of their respective theories would not support Copernicus's over Ptolemy's, despite the falsity of the latter. Ptolemy's astronomy was the first ever unification of physics and astronomy. While Aristotle's physics was then the leading account of motion of objects on Earth, it lacked a detailed astronomical theory. Ptolemy's astronomy addressed this issue, but since it was built on Aristotle's physics, it faced a serious difficulty at once: the observed motion of Mars included a retrograde trajectory that seemed to conflict with the Aristotelian requirement that planets move in circles. Undeterred, Ptolemy introduced an important conceptual innovation: the epicycle, a circle centered on the circumference of another circle. Epicycles allowed him to explain how the motion of Mars could be retrograde. Ptolemy's great achievement was an astronomical theory that reasonably fitted the celestial data, while being consistent with the dominant physics of his time. The resulting theory was unified (it brought together physics and astronomy), explanatory (it systematically accounted for the motion of planets as orbiting in circles around a stationary Earth, thus preserving appearances), and simple (since all celestial motion was circular).

In contrast, Copernicus' theory compares poorly on these virtues. First, it is less unified than Ptolemy's. At the time in which Copernicus developed his conception, there was no physics of a moving Earth. In the first chapter of *On the Revolutions of the Heavenly Spheres* (Copernicus, 1543/1976), Copernicus had to address well-known objections to the very possibility of a moving Earth, which involved concerns of the kind: If the Earth is moving, how can birds return to their nests or why does a dropped stone end up at one's feet rather than a few inches away? Not having a physics of a moving Earth at his disposal, Copernicus tried to deflate the objections as swiftly as he could, but it was very clear that the lack of a proper physics raised substantial difficulties for his entire project. In proposing an heliocentric system in the absence of a suitable physics, Copernicus clashed with the highly unified conception that Ptolemy's theory provided. On the unification front, Copernicus clearly lost.

It is often claimed that Copernicus' theory is simpler than Ptolemy's, since the former does not invoke epicycles, but not even this is right. Although epicycles are unnecessary to account for Mars' retrograde motion in Copernicus' theory, they are still required to explain other planetary motions (Kuhn, 1957; Cohen, 1983). The fact that the entire planetary system needed to be reconfigured in a way that went against the best physical theory of the time—namely, Aristotle's physics—also counted against the theory's simplicity. It required a dramatic shift from the then-accepted conception of the universe in no small part by moving away from a highly intuitive foundational physics.

The explanatory power of Copernicus's theory in comparison with Ptolemy's was also compromised. The former accounts for the phenomena on the assumption that the Earth is moving around the Sun rather than the other way around. At the time in which Copernicus' theory was formulated, however, there was surprisingly little evidence for that assumption. It was one of the main accomplishments of Galileo to valiantly forge considerations in support of the Copernican theory, in works such as *Starry Messenger* (Galilei, 1610/2009) and *Dialogue Concerning the Two Chief World Systems: Ptolemaic and Copernican* (Galilei, 1632/1953). Given the natural tendency to let predecessors of modern, more correct theories inherit illicitly the glow of plausibility from their

successors, it must be fully appreciated that Galileo's arguments flew in the face of the best theory of his time. That Venus has phases similar to the Moon, that Jupiter has satellites similar to the Earth, that the Moon has valleys and mountains not unlike the Earth's were claims he made on the basis of interpreting data acquired by using an instrument (the telescope), the reliability of which regarding celestial matters, according the Aristotelian theory, had no basis. After all, on that theory, the celestial domain, which functions immutably in accordance with perfect circles, operates in an entirely different way than does the sublunar domain, which is subject to change, decay and erratic motion. With hindsight, we now judge that Galileo was factually correct and right to defend the Copernican conception. At the time, whatever intellectually respectable grounds there might have been for choosing the Copernican theory over the Ptolemaic, they cannot have been that Copernicus's theory was more explanatory than Ptolemy's. Quite to the contrary, the available evidence and the then-accepted physics went against the Copernican conception not least because it undermined the evidential value of the instrument that was crucial for gathering Galileo's supporting evidence (Feyerabend, 1975). A serious application of theoretical utility considerations in Galileo's day would have favoured Ptolemy's theory over Copernicus's, not the other way around.

Only when Newton formulated his theory in *Philosophia Naturalis Principia Mathematica* (Newton, 1687) was a unification between physics and astronomy finally obtained again (Cohen, 1983). Newton's theory embodies all of the commonly-cited theoretical virtues. It is simple (it relies on gravity); it is unified (it brings together motion of celestial bodies and objects on the surface of the Earth, it accounts for the tides and other natural phenomena), and it is explanatory (it explains, in light of gravity, why objects, whether sub- or superlunar, move the way they do).

Despite this impressive feat regarding theoretical virtues, Newton's theory turns out to be false. There is no Newtonian action-at-a-distance gravitational force, though; if relativity theory is right, there is curvature of space-time. If what was perhaps the most theoretically virtuous physical theory turns out to have been false, it is hard to see why being virtuous could be properly treated as a mark of truth.

Furthermore, such virtue is not even necessary for the truth. The shift to (non-relativistic) quantum mechanics cannot be properly accounted for by invoking a theoretical utility argument. After all, (non-relativistic) quantum mechanics does not satisfy the theoretical virtues, despite the impressive empirical success of the theory. The theory is not simple, since it is highly counterintuitive, and requires a number of distinctive mathematical procedures. It introduced a new disunity into physics, since it is inconsistent with relativity theory. Even though it successfully predicted a number of new phenomena, its explanatory power is suspect, given its multiple, incompatible interpretations. Those incompatible interpretations leave us without a unique account of the underlying realities that give rise to the observable phenomena. It is unclear that the theory has fully explained those phenomena given that—in light of admissible interpretations—it cannot be determined whether the outcome of observations emerges because upon measurement a physical system evolves from a superposition of states or because the world splits into different worlds, among many possibilities (Hughes, 1989; van Fraassen, 1991; Wallace, 2012). If the underlying features of the system are not well understood, the explanatory capacities of the relevant theory are not driving its adoption. If the commonly-cited theoretical virtues really are markers of truth, they seem not to be driving crucial aspects of scientific activity and, so, the roles those virtues play in the sciences are not good guides to good method for metaphysicians.

The situation is no different in a non-empirical domain. Frege's original formulation of arithmetic offers a clear example. Frege's is a theory that embodies all theoretical virtues: it is simple (it relies on logic and definitions); it is unified (it brings together arithmetic and logic—in fact, for the logicist, arithmetic is nothing but logic plus definitions); it has expressive power (arithmetical truths can all be expressed in terms of suitable second-order statements); and it is explanatory (it accounts for the nature of numbers while preserving all arithmetical truths). Sadly, the theory is also false. Worse than being merely false, Frege's original formulation of arithmetic is not even consistent. Thus, the satisfaction of theoretical virtues is neither necessary nor sufficient for a theory to be truth conducive.

Copernican theory, Newtonian mechanics, and quantum mechanics are all unquestionably extremely successful theories. That arguments based on theoretical virtues fail to ratify these theories for what they are correctly taken to be—namely, clear cases of successful scientific theories—undermines any claim that such arguments are so central to actual scientific methodology. Metaphysicians assume an inaccurate account of scientific practice.

Suppose, though, that some scientists do reason from theoretical virtue to theoretical accuracy when empirical adequacy does not suffice. Those scientists owe us, then, exactly the same non-question-begging arguments owed to us by metaphysicians. Both enterprises purport to discover or determine what there is and how it is. It is easy to understand the pragmatic value of theories with the typically-sought virtues, being easier to work with as they are. It is also easy to see how they are proxy for other things we value. Theories that are simpler, unified, powerfully expressive, and explanatory may possess non-theoretical virtues that we value. That virtuous theories better track reality than do their less virtuous competitors is established by neither of these facts. First, these facts alone are not enough to make theoretical virtues truth-conducive. Second, it is similarly inadequate to build into a given theory one's preferred theoretical virtue (such as, simplicity) and then invoke the satisfaction of such a virtue as a sign of the theory's accuracy—unless one has independent reasons to believe that reality has the virtue in question (in this case, that it is indeed simple). But whether reality has or has not the corresponding virtue is precisely what one is trying to determine in the first place; one cannot invoke a theoretical utility argument to determine whether the theoretical virtues in question obtain without begging the question. In other words, once both the scientific and metaphysical enterprises are taken to be in the business of describing important features of reality, it is manifestly inadmissible to build values of reality's parameters into the characteristics of theories we seek and then use those characteristics as signs of theoretical accuracy. That is for both scientists and metaphysicians not to do their respective jobs.

4.2 A Philosophical Variation

In metaphysics, the use of theoretical utility considerations in support of one's ontology has been strongly advocated by David Lewis. Drawing on an analogy with set theory, he notes:

> Set theory offers the mathematician great economy of primitives and premises, in return for accepting rather a lot of entities [...]. The price is right; the benefits in theoretical unity and economy are well worth the entities.

> [...] If we want the theoretical benefits that talk of *possibilia* brings, the most straightforward way to gain honest title to them is to accept such talk as the literal truth. It is my view that the price is right, if less spectacularly so than in the mathematical parallel. The benefits are worth their ontological cost. Modal realism is fruitful; that gives us good reason to believe that it is true (Lewis, 1986, p. 4).

Why think that theoretical benefits provide reason to think that the relevant talk is literally true? We do not usually take being in our interest to be a sign of truth. What, exactly, is the link between utility and truth conduciveness? This link is assumed by Lewis and others, but its assumption does not suffice, since it is easy to multiply examples of benefits and convenience conferred by falsehoods.

Strictly speaking, Lewis urges on us a cost-benefit analysis. How odd to employ something appropriate for a decision context (where deliberation is directed at what to do) in an epistemic context (where deliberation is directed at what to believe). Grounds for deciding what to do are not grounds for believing what is the case, regardless of how much what one believes enters into one's deliberations about what to do. One can decide to act against the facts, contrary to evidence, and in conflict with one's best judgement. Actions toward which deliberations are directed are typically voluntary in ways beliefs usually are not. One cannot simply decide to believe that there is a pink elephant in the room, even though one can decide to act as though there were. Because decision contexts are fundamentally different from epistemic context, cost-benefit analyses are fundamentally inapt for metaphysical inquiry precisely because they inevitably conflate pragmatic and epistemic reasons. For the "costs" and "benefits" of a theory to have epistemic merit the relevant features that are transgressed by the costs and to which the benefits conform must have been either assumed to be characteristics of reality or to have been justified by other means. If assumed, questions about the structure of reality have been begged. If justified by other means, the argument from theoretical virtue is redundant for establishing that characteristic of the world.

After the cost-benefit analysis is performed, one may end up with reasons to *accept* a metaphysical theory, since it has a number of significant benefits without too many costs. By accepting such a theory on the basis of theoretical virtues, one would have pragmatic reasons to prefer working with it: simplicity, expressive and explanatory power are all good theoretical traits; they provide at least some of what one wants from a theory. As argued above, however, they fail to offer what is needed to capture the basic structures of reality: they are not truth conducive.

In light of these considerations, the first premise of the theoretical virtues argument does not go through. We now move to the second.

5 The Virtue Premise

A key feature of the argument from theoretical virtues is to establish that the theories under consideration do exhibit the relevant virtues. This is the role of the second premise of the argument. For some of the virtues, however, in order to claim that a given theory is virtuous and, indeed, more virtuous than its rivals, one must suppose that the theory in question is true; otherwise, the virtues in question do not obtain. To support this point from the sciences, we consider Newtonian physics and Mercury's perihelion. From metaphysics, we consider modal realism.

Mercury's orbit of Mercury did not conform to predictions based on Newtonian physics. Physicists posited the existence of a planet—Vulcan—between Mercury and the

Sun whose presence would account for the deviation that was observed in Mercury's orbit when it gets closer to the Sun. After multiple unsuccessful attempts to detect the planet and the assessment that observational capabilities sufficed to conclude that if Vulcan existed and explained Mercury's orbital anomalies, Vulcan would have been observed, physicists concluded that it does not exist. Thus, Newtonian mechanics has no explanatory power regarding fine-grained details of Mercury's orbit. All claims to explanatory virtues depend on the correctness of the theory and cannot be used to establish its correctness. Thus exposed is Williamson's use of the explanatory virtue as a mark of truth while characterizing a theory as having that virtue in modal terms, i.e., that it would explain relevant matters better, if true (Williamson, 2016, p. 266). "We should think it true, because it would do a better job of explaining things were it true" is an expression of hope, rather than an articulation of a plausible basis for thinking a theory true. That it *would do* a better job is thought to justify that it *does* a better job. That it *does* a better job is thought to justify embracing it as true. A conclusion that a theory is true, however, cannot be warranted on the basis of a premise regarding its potential to do or to be something or other.

A problem specific to any claims to explanatory virtue parallels claims to simplicity made above. Making breadth or depth of explanation a virtue for identifying a theory as (more likely to be) correct is to assume that the world is not intractably complex, disjoint, disconnected, and unpredictable. This may be the working assumption of any scientific enterprise that is at all theoretical and not limited to descriptive activities. That it must be presupposed for rational activity does not provide it any epistemic force at all. The mutually inconsistent, unintuitive interpretations of quantum theory have forced us to recognize that assumptions about the availability of satisfying explanations is, indeed, an assumption. To that degree, we have grounds for recognizing that the world may not conform to our explanatory assumptions, showing the assumption to be illicit when taken as a marker of truth.

Turning to metaphysics, we find a similar fault in the work of David Lewis, who has engaged in a rather comprehensive project of defending his well-known modal realism (Lewis, 1986). According to this view, modal vocabulary and identity are exchanged for quantification over spatiotemporally disconnected spatiotemporal wholes and counterpart relations. Much effort has been expended both by Lewis and sympathizers to work out the subtle details, which we do not dispute. The concern is that it is only by assuming the truth of modal realism that one can obtain any explanatory advantage from the view. However, whether the view is true or not is precisely what needs to be determined. In the end, to obtain support for the second premise of the theoretical utility argument, one needs to assume the truth of the very theory that argument is meant to support.

Consider that, no doubt, Lewis could have been a dentist rather than a philosopher. This truth, however, ends as a falsehood if there is no plurality of worlds. Similarly, in the absence of that plurality, the modal realist has no resources to account for the objectivity of the fact that Lewis could have been a dentist, and thus argue that modal realism provides a better account of modal reality than its rivals. Affirming the virtue premise of any argument from theoretical virtues is a dialectically illicit affirmation. The virtue accrues to a theory only if it is true, but the argument depending on the theory's virtue is precisely an argument *for* the theory's truth.

Similarly for claims that counterpart relations will do when compared to identity across possible worlds. The table before us is breakable; this very table is. What does that come to? If there is only one spatiotemporal whole, then it cannot come to there being another in which a numerically distinct table represents our table sufficiently

well that the breaking of the other suffices to make it that ours could have broken. This is not a rehashing of Kripke's "Humphrey" objection (Kripke, 1980). We do not complain that we are interested in whether *this* table could have broken, but are treated only to a story about how *that* table did break. We do not impose here a constraint on the semantics and metaphysics of modality that rules out modal realism at the outset. We complain only that the virtue of exposing the reality underlying the possibility of our table breaking accrues to modal realism only if modal realism is correct, making a claim to its virtue illicit in an argument for its truth.

These considerations highlight that using theoretical virtues as tools of persuasion is ill conceived. They require the commitment to what one is trying to establish in the first place. In the end, this puts pressure on the very idea of using theoretical utility arguments as a source of epistemic justification.

We have now exposed why the more traditional way of understanding abductive inference is not useful when thinking about permissible philosophical argumentation. In its general structure, the argument has the form of affirming the consequent, where the conditional premise has as its antecedent a theory T and its consequent e, the relevant evidence. In empirical contexts, we have the advantage that, in properly chosen conditions, e is determinable independent of a prior determination of the correctness of T. The very features of metaphysical inquiry that drive philosophers to theoretical utility arguments are those that prevent the acceptability of these arguments as legitimate tools of persuasion about the theory in question precisely because this kind of independent determination of relevant facts is impossible. Hardly anything about whether a theory really possesses the virtues cited on its behalf is available absent assuming the theory. Further complaints about whether any versions of affirming the consequent provides non-deductive grounds for an empirical theory are best left for another occasion. It suffices here to note that no such version is available to philosophers for the reasons we have given.

6 Conclusion

Theoretical utility arguments have been prominent in the support of the truth of metaphysical theories. We argued that both in science and in metaphysics such arguments fail to establish the truth of the theories in question.

Empiricism was long characterized by a distinction between evidence and theory. Critics, such as Quine, maintained that this distinction is untenable and opted for confirmational holism. Theories stand or fall as wholes and no part is immune from revision (Quine, 1951). Consequently, any theoretical or observational claim can be retained so long as suitable adjustments are made to other factual or theoretical statements. Absent any possible appeal to evidence, holists depend on assessing the theoretical virtues of competing theories. A significant consequence of what we have argued here is that such assessments are dead ends for confirmational holists who are realists about theories so chosen. Those who wish to retain their holism and their arguments from theoretical virtues must adopt some non-realist attitude toward their chosen theories.

Of course, metaphysicians could devise alternative strategies to uncover the truth about the world. It is their burden, then, to establish that such strategies—whatever they might be—tend toward the truth. Given the nature of the metaphysical enterprise, it is not at all clear how this is to be accomplished. Nevertheless, those wishing for robust metaphysical results must face the task at hand.

References

Armstrong, D. M. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.

—. 1997. *A World of States of Affairs.* Cambridge: Cambridge University Press.

Bigelow, John. 2010. "Quine, Mereology, and Inference to the Best Explanation." *Logique et Analyse* 53:465–482.

Bigelow, John and Pargetter, Robert. 1990. *Science and Necessity.* Cambridge: Cambridge University Press.

Biggs, Stephen. 2011. "Abduction and Modality." *Philosophy and Phenomenological Research* 83:283–326.

Bueno, Otávio and Shalkowski, Scott A. 2015. "Modalism and Theoretical Virtues: Toward an Epistemology of Modality." *Philosophical Studies* 172:671–689.

Cohen, I. Bernard. 1983. *The Birth of a New Physics.* New York: Norton and Company.

Colyvan, Mark. 2006. "Scientific Realism and Mathematical Nominalism: A Marriage Made in Hell." In Colin Cheyne and John Worral (eds.), *Rationality and Reality: Conversations with Alan Musgrave*, 225–37. Dordrecht: Springer.

Copernicus, Nicholas (ed.). 1543/1976. *Copernicus, on the revolutions of the heavenly spheres: a new translation from the Latin.* Newton Abbott: David and Charles. With an introduction and notes by A. M. Duncan.

Dorato, Mauro. 2012. "Mathematical Biology and the Existence of Biological Laws." In Dennis Dieks, Wenceslao J. Gonzalez, Stephan Hartmann, Michael Stöltzner, and Marcel Weber (eds.), *Probabilities, Laws, and Structures*, 109–121. Dordrecht: Springer.

Ellis, Brian. 2009. *The Metaphysics of Scientific Realism.* Chesham: Acumen.

Feyerabend, Paul. 1975. *Against Method.* London: New Left Books.

Fischer, Bob. 2017. *Modal Justification via Theories.* Dordrecht: Springer.

Galilei, Galileo. 1610/2009. *Galileo's Sidereus Nuncius, or A Sidereal Message.* Sagamore Beach, MA: Science History Publications. Translated from the Latin by William R. Shea; introduction and notes by William R. Shea and Tiziana Bascelli.

—. 1632/1953. *Dialogue Concerning the Two Chief World Systems - Ptolemaic & Copernican.* Sagamore Beach, MA: Science History Publications. Translated by Stillman Drake; foreword by Albert Einstein.

Hughes, R. I. G. 1989. *The Structure and Interpretation of Quantum Mechanics.* Cambridge, MA: Harvard University Press.

Kripke, Saul. 1980. *Naming and Necessity.* Cambridge, MA: Harvard University Press.

Kuhn, Thomas. 1957. *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought.* Cambridge, MA: Harvard University Press.

Lewis, David. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

McLaughlin, Brian P. 2010. "Consciousness, Type Physicalism, and Inference to the Best Explanation." *Philosophical Issues* 20:266–304.

Newton, Isaac. 1687. *Philosophiæ Naturalis Principia Mathematica*. London: Jussu Societas Regiæ.

Paul, L. A. 2012. "Metaphysics as Modeling: The Handmaiden's Tale." *Philosophical Studies* 160:1–29.

Psillos, Stathis. 2005. "Scientific Realism and Metaphysics." *Ratio* 18:385–404.

Quine, W. V. O. 1951. "Two Dogmas of Empiricism." *Philosophical Review* 60:20–43.

Saatsi, Juha. 2017. "Explanation and Explanationism in Science and Metaphysics." In Matthew Slater and Zanja Yudell (eds.), *Metaphysics and the Philosophy of Science: New Essays*, 163–191. New York: Oxford University Press.

Sider, Theodore. 2011. *Writing the Book of the World*. Oxford: Clarendon Press.

Sider, Theodore, Hawthorne, John, and Zimmerman, Dean (eds.). 2008. *Contemporary Debates in Metaphysics*. Oxford: Blackwell.

Swoyer, Chris. 1999. "How Ontology Might Be Possible: Explanation and Inference in Metaphysics." *Midwest Studies In Philosophy* 23:100–131. doi:10.1111/1475-4975.00006.

—. 2008. "Abstract Entities." In Theodore Sider, John Hawthorne, and Dean W. Zimmerman (eds.), *Contemporary Debates in Metaphysics*, 11–31. Malden, MA: Blackwell.

van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Clarendon Press.

—. 1991. *Quantum Mechanics: An Empiricist View*. Oxford: Clarendon Press.

Wallace, David. 2012. *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. New York: Oxford University Press.

Williamson, Timothy. 2016. "Abductive Philosophy." *The Philosophical Forum* 47:263–280.