

Systematizing the theoretical virtues

Michael N. Keas¹ 

Received: 14 April 2016 / Accepted: 22 February 2017 / Published online: 10 March 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract There are at least twelve major virtues of good theories: evidential accuracy, causal adequacy, explanatory depth, internal consistency, internal coherence, universal coherence, beauty, simplicity, unification, durability, fruitfulness, and applicability. These virtues are best classified into four classes: evidential, coherential, aesthetic, and diachronic. Each virtue class contains at least three virtues that sequentially follow a repeating pattern of progressive disclosure and expansion. Systematizing the theoretical virtues in this manner clarifies each virtue and suggests how they might have a coordinated and cumulative role in theory formation and evaluation across the disciplines—with allowance for discipline specific modification. An informal and flexible logic of theory choice is in the making here. Evidential accuracy (empirical fit), according to my systematization, is not a largely isolated trait of good theories, as some (realists and antirealists) have made it out to be. Rather, it bears multifaceted relationships, constituting significant epistemic entanglements, with other theoretical virtues.

Keywords Theoretical virtues · Inference to the best explanation · Epistemic value · Aesthetics · Prediction · Science–technology relations

1 Introduction

Theoretical virtues are the traits of a theory that show it is probably true or worth accepting. Although the identification, characterization, classification, and epistemic

✉ Michael N. Keas
mkeas@icloud.com

¹ History and Philosophy of Science, Faculty of Humanities, The College at Southwestern,
PO Box 22206, Fort Worth, TX 76122, USA

standing of theory virtues are debated by philosophers and by participants in specific theoretical disputes, many scholars agree that these virtues help us to infer which rival theory is the best explanation (Lipton 2004). Analysis of widely accepted theories, especially in the natural sciences, can help us to more skillfully use these tools in all disciplines. I offer a new systematization of the theoretical virtues to deepen our understanding of them.

The most widely accepted theories across the academic disciplines usually exhibit many of the same theoretical virtues listed below. Each virtue class contains at least three virtues that sequentially follow a repeating pattern of progressive disclosure and expansion. This pattern, and other systematic features, will become apparent as we explore the theoretical virtues.

Evidential virtues

1. *Evidential accuracy*: A theory (T) fits the empirical evidence well (regardless of causal claims).
2. *Causal adequacy*: T's causal factors plausibly produce the effects (evidence) in need of explanation.
3. *Explanatory depth*: T excels in causal history depth or in other depth measures such as the range of counterfactual questions that its law-like generalizations answer regarding the item being explained.

Coherential virtues

4. *Internal consistency*: T's components are not contradictory.
5. *Internal coherence*: T's components are coordinated into an intuitively plausible whole; T lacks ad hoc hypotheses—theoretical components merely tacked on to solve isolated problems.
6. *Universal coherence*: T sits well with (or is not obviously contrary to) other warranted beliefs.

Aesthetic virtues

7. *Beauty*: T evokes aesthetic pleasure in properly functioning and sufficiently informed persons.¹
8. *Simplicity*: T explains the *same facts* as rivals, but with *less* theoretical content.
9. *Unification*: T explains *more kinds of facts* than rivals with the *same* amount of theoretical content.

Diachronic virtues

10. *Durability*: T has survived testing by successful prediction or plausible accommodation of new data.
11. *Fruitfulness*: T has generated additional discovery by means such as successful novel prediction, unification,² and non ad hoc theoretical elaboration.

¹ Section 5.1 considers how alternative accounts of aesthetics might bear upon my taxonomy.

² Sections 5.2 and 6.2.1 distinguish “unification” as a kind of aesthetic virtue and as a mode of diachronic fruitfulness.

12. *Applicability*: T has guided strategic action or control, such as in science-based technology.

I will show how this theoretical virtue taxonomy is more illuminating than others (e.g., it helped me identify the virtue of applicability, which is absent from earlier lists and taxonomies). Previous attempts at understanding and classifying the theoretical virtues would have been more successful had they attended to all of these major virtues and their relations. After a glimpse at early attempts to understand and classify the theoretical virtues, we will explore each of the virtues and their taxonomic relations.

2 Early attempts to understand and systematize the theoretical virtues

In response to critics who charged him with undermining science's rationality, Kuhn wrote:

I have implicitly assumed that, whatever their initial source, the criteria or values deployed in theory-choice are fixed once and for all, unaffected by their transition from one theory to another. Roughly speaking, but only roughly speaking, I take that to be the case. If the list of relevant values be kept short ... and if their specification be left vague, then such values ... are permanent attributes of science (1977, p. 335).

Kuhn listed five such permanent values (theoretical virtues) operative in theory choice: accuracy, consistency, scope (unification), simplicity, and fruitfulness (p. 332). He persisted in this understanding during his final years of work, affirming that the theoretical virtues are “necessarily permanent, for abandoning them would be abandoning science” (Kuhn 1993, p. 338). Neither Kuhn, nor most others contributing to this topic, have proposed a classification system that indicates how the different theoretical virtues relate to each other.

Laudan (1984) ventured slightly into theory virtue systematization (only at a proposed taxonomic level *above* my four “classes”) by arguing that most theoretical virtues in scientific practice fail to possess strict “epistemic” credentials as traditionally conceived in analytic epistemology. Rather, he classified all theory virtues within the larger category of “cognitive virtues or values,” of which “the epistemic virtues form a proper ... subset” (2004, p. 19). Many philosophers of science use the term “epistemic” in a broad sense that often approximates Laudan's “cognitive” when talking about theory virtues. For example, Reiss and Sprenger (2014) cite Laudan in this regard and recommend this broader sense of “epistemic” (as do I):

Sometimes epistemic values are regarded as a subset of cognitive values and identified with values such as empirical adequacy and internal consistency that directly bear on the veracity of a scientific theory (Laudan 2004). Values such as scope and explanatory power would then count as cognitive values that express scientific desiderata, but without properly epistemic implications. We have decided, however, to adopt a broader reading of “epistemic” where truth is not the only aim of scientific inquiry, but supplemented by providing causal

mechanisms, finding natural laws, creating understanding, etc. In this sense, values such as scope or explanatory power contribute to achieving our epistemic goals. Neat distinctions between strictly truth-conducive and purely cognitive scientific values are hard to come by (see [Douglas 2013](#) for a classification attempt).

We will glean important insights from Douglas' classification system later in the present study. For now, we must finish surveying early attempts to understand theoretical virtues. Many of the leading twentieth century virtue analysts emphasized the integration of the history and philosophy of science. Besides Thomas Kuhn and Larry Laudan, Ernan McMullin fits prominently within this early HPS generation of theory virtue thinkers.

[McMullin \(1983\)](#) was a pioneer of the broad sense of “epistemic” (in contrast to Laudan's narrow sense) in regard to theory virtues. He also proposed the first taxonomy of theoretical virtues ([1996, 2014](#)), which grew out of his earlier work on the nature of theory virtues, back when he called them “epistemic values” ([1983, 1987, 1993](#)). In his final theory virtue systematization ([2014](#)), evidential accuracy is number one on his list, but he calls it “empirical fit.” This is, says McMullin, “the primary theory virtue” because “the first requirement of theory is to account for data already in hand” ([2014](#), pp. 563–564). For McMullin, this theory trait is one of a kind. He classifies the remaining theoretical virtues into three groups of complementary virtues: internal, contextual, and diachronic. These virtues complement how well a theory fits the evidence, he argues.

How does McMullin's taxonomy compare with mine? His two internal theoretical virtues, internal consistency and internal coherence, are identical to the first two of my coherential virtues. My third coherential virtue, universal coherence, is roughly equivalent to McMullin's second category: contextual virtues (he includes here “external consistency” with related knowledge and consistency with deeper metaphysical foundations). He dismisses the aesthetic virtues (focusing his criticism on simplicity) as too problematic to be worthy of inclusion, and then argues that the remaining theoretical virtues fall within the category of “diachronic.” This third category consists of virtues that require time to show a track record of performance after a theory's origin. He includes here durability, fertility (roughly identical to my “fruitfulness”), and consilience (which is the “unification” mode of fruitfulness in my system). My taxonomy, in contrast to McMullin's, characterizes unification as belonging primarily to the aesthetic class of theoretical virtues, and only secondarily is depicted as a component of diachronic fruitfulness.

While my theoretical virtue systematization borrows from McMullin's work, it goes beyond his accomplishment primarily in how I develop the following: (i) the evidential virtues, which McMullin reduces to the one-of-a-kind “empirical fit” virtue, and does not treat the closely related evidential virtues of causal adequacy and explanatory depth, and (ii) the aesthetic virtues, which McMullin dismisses as irrelevant factors in rational theory choice. I will unpack these improvements, and more, below.

3 Evidential theoretical virtues: evidential accuracy, causal adequacy, and explanatory depth

The evidential virtues indicate different facets of how well a theory accounts for the entities, events, and regularities in the world. This class of virtues is best comprehended through historical case studies.

3.1 Evidential accuracy (TV1)

Evidential accuracy is instantiated in a theory when it fits the empirical evidence well. Prior to Galileo's telescopic discoveries of 1609–1610, the evidential accuracy of the contending geocentric and heliocentric astronomical systems were roughly equal. After the telescope showed Venus going through many moon-like phases that would be impossible if Ptolemy's ancient geocentric theory were true, this narrowed the contending systems down to Copernicus' heliocentric theory and Tycho Brahe's geoheliocentric system. In the Tychonic system the Sun and fixed stars revolve around a central-stationary earth. The planets revolve around the sun as their (moving) center. If evidential accuracy were the only recognized criterion for theory choice at this time, then astronomers would have had insufficient reason to accept the heliocentric theory. Thankfully, multiple virtuous criteria for theory choice were available, especially as articulated by Kepler ([Jardine and Kepler 1984](#)). In addition to evidential accuracy—conceived since antiquity mainly as the match between mathematical astronomical theory (chiefly combinations of circular motions) and the observed apparent motions of planets—Kepler also insisted that astronomers evaluate theories by the physical plausibility of causal components.

3.2 Causal adequacy (TV2)

A theory is causally adequate if it specifies causal factors that plausibly produce the effects in need of explanation. [Longino \(1996\)](#) argues that often an adequate causal account must include multiple factors operating in a web of mutual interaction, rather than a single unidirectional cause-and-effect relationship. In some cases a good explanation includes appeals to geometrical or other reasons. For example, one might explain a free particle's path by its conformity to a geodesic in spacetime—a geometrical spacetime reason ([Nerlich 1979](#)). Such an explanation might be “causally adequate” in a broad sense.

Robustness³ analysis in comparative modeling can help establish causal adequacy ([Knuuttila and Loettgers 2011](#)), as Weisberg explains.

The key comes in ensuring that a sufficiently heterogeneous set of situations is covered in the set of models subjected to robustness analysis. If a sufficiently heterogeneous set of models for a phenomenon all have the common structure, then it is very likely that the real-world phenomenon has a corresponding causal

³ See [Soler et al. \(2012\)](#) for other senses of “robustness” that figure into other theoretical virtues.

structure. This would allow us to infer that when we observe the robust property in a real system, then it is likely that the core [causal] structure is present and that it is giving rise to the property (2006, p. 739).

A celebrated episode in the history of geology illuminates the relation between evidential accuracy and causal adequacy. During much of the twentieth century the theory of continental drift enjoyed a modest degree of evidential accuracy (e.g., geological continuity between continents, including complementary shapes and similar fossils), but lacked causal adequacy. Soon after he announced the theory, Alfred Wegener urged patience regarding the theory's causal inadequacy:

The question concerning which forces cause the horizontal displacement of the continents we have advocated is so obvious that I cannot completely overlook it, although I am of the opinion that it is premature. It is unquestionably necessary first to establish exactly the reality and the type of the displacements before one can hope to fathom their causes (Greene 2015, p. 264).

After a few decades of being underappreciated, the theory gained extensive new support from paleomagnetic and related seafloor spreading studies. This vastly expanded the theory's evidential accuracy. Ironically, despite the earlier decades of complaints of a lack of a plausible causal mechanism for continental drift, the updated theory (known as plate tectonics) became widely accepted in the late 1960s based on its stronger evidential accuracy, despite its continued lack of causal adequacy (Frankel 2012). Later evidence implicated circulating convection currents in the hot soft mantle below the relatively rigid oceanic and continental plates. Many argued that this explained plate movement, rendering the theory causal adequate.

A third kind of evidential virtue, explanatory depth, increased as the theory of mantle convection currents was supplemented with several other associated factors. Geophysicists found other processes that contribute to plate movement, which generated more causal history depth.

3.3 Explanatory depth (TV3)

A theory exhibits explanatory depth when it excels in causal history depth or in other depth measures such as the range of counterfactual questions that its law-like generalizations answer regarding the item being explained. Causal history depth is often characterized in a causal-mechanical way by how far back in a linear or branching causal chain one is able to go (or perhaps we encounter a mutually interacting web of causal factors that outstrips mechanical explanation). While causal adequacy (TV2) is about basic causal-explanatory satisfaction (going from effect back to immediate putative cause or causes), causal history depth (an instance of TV3) goes back further.

Explanatory depth comes in at least two varieties, depending upon whether it pertains primarily to events or laws. Causal history depth focuses on events. The second main variety of explanatory depth, which is law-focused, has been elucidated best in Hitchcock and Woodward's causal-counterfactual account of explanation, which is based on a particular sort of "range" or generality that is fundamentally different from unification (TV9 also known as broad scope; see Sect. 5.2). Unification refers to

“generality with respect to objects or systems other than the one that is the focus of explanation.” Hitchcock–Woodward explanatory depth, by contrast, targets “range” in the sense of “generality with respect to *other possible properties of the very object or system that is the focus of explanation*” (2003, p. 182).

Newton’s account of free fall possessed more explanatory depth than Galileo’s. Newton explained not just free fall very near earth’s surface (the restricted range of Galileo’s theory), but also free fall toward earth starting from any distance. Furthermore Newton could explain free fall toward a hypothetically “altered earth”—perhaps if there is a change in its mass and radius, or if one works with another planet or a star that has such an alternative mass and radius. So the Newtonian explanation of free fall remains invariant through a larger range of investigator interventions. In short, Newton’s “free fall” account is explanatorily deeper than Galileo’s because it handles a larger range of counterfactual (what-if-things-had-been-different) questions about the *same kind of phenomena* (free fall in various circumstances). Of course, Newton’s physics (three laws of motion and universal gravitation) explains many *other kinds* of phenomena beyond free fall about which Galileo was silent, but that constitutes the taxonomically disparate theoretical virtue of unification (TV9), not explanatory depth (TV3).

Hitchcock and Woodward pinpoint explanatory depth as illustrated by Galileo’s and Newton’s “free fall” this way:

In these sorts of cases, claims about the invariance of a relationship under changes in *background conditions* are transformed into claims about invariance under interventions on *variables figuring in the relationship* through the device of explicitly incorporating additional variables into the relationship. For example, an intervention that increases the mass of the earth would count as an intervention on *background conditions* with respect to Galileo’s law, but as an intervention on a *variable explicitly figuring in* Newton’s laws. This is, perhaps, the most fundamental way in which one generalization can provide a deeper explanation than another (2003, p. 188).

Notice the progression within the evidential theoretical virtues from achieving a basic evidential fit (evidential accuracy), to identifying a minimally complete causal story (causal adequacy), and finally to deepening the explanation of the evidence in either an event-event or law-like way. Each additional virtue builds upon the previous ones, with evidential accuracy at the foundation. Typically causal adequacy requires evidential accuracy and explanatory depth requires causal adequacy. As hinted above, these evidential theoretical virtues are better grasped in the light of the leading theories of what it means to “explain” something.

3.4 Evidential theoretical virtues in the light of theories of explanation

One need not exclusively endorse any one of the three leading accounts of explanation—causal–mechanical, causal–counterfactual or unificationist views—in order to glean insight from this discussion to better understand several of the theoretical virtues. The causal–mechanical theory of explanation suggests that one explains

something by identifying the causal mechanisms that generated it. This might involve identifying a single cause, or a complicated causal history. This view of explanation helps motivate recognizing (i) causal adequacy as an evidential theoretical virtue and (ii) the detection of deep causal history as one measure of the evidential virtue of explanatory depth.

The causal-counterfactual view of explanation provides the basis for yet another measure of explanatory depth, namely, the range of counterfactual questions that a theory's law-like generalizations answer regarding the item being explained. According to this account "to explain" is to answer what-if-things-had-been-different questions, ideally by means of experimental control in which all known relevant factors are held constant except for one factor that the investigator manipulates to see what will happen.

Hitchcock and Woodward offer a clear statement of this theory of explanation and its depth dimension.⁴

One generalization can provide a deeper explanation than another if it provides the resources for answering a greater range of what-if-things-had-been-different questions, or equivalently, if it is invariant under a wider range of interventions. That is, generalizations provide deeper explanations when they are more general. It is important, however, to understand generality in the right way: generality with respect to hypothetical changes in the system at hand. By focussing on the wrong sort of generality—generality with respect to systems other than the one whose features are to be explained—rival accounts of explanation such as Hempel's D-N model and Kitcher's unificationist theory have been unable to provide adequate accounts of explanatory depth (2003, p. 198).

Kitcher outlines a third major theory of explanation: the unificationist view.

I have sketched an account of explanation as unification, attempting to show that such an account has the resources to provide insight into episodes in the history of science and to overcome some traditional problems for the covering law model. In conclusion, let me indicate very briefly how my view of explanation as unification suggests how scientific explanation yields understanding. By using a few patterns of argument in the derivation of many beliefs we minimize the number of types of premises we must take as underived. That is, we reduce, in

⁴ Much of my account of explanatory depth draws from Hitchcock and Woodward (2003), who distinguish their approach from Kitcher's unificationist account and Lewis' causal account. Ylikoski and Kuorikoski (2010) extend the Hitchcock–Woodward theory, and explicate a cluster of theoretical virtues in its light. They focus on "explanatory power" and argue that it has five components (none of which are aesthetic or diachronic virtues): non-sensitivity, precision, factual accuracy, degree of integration and cognitive salience. Their non-sensitivity is roughly my explanatory depth, their precision and factual accuracy are other considerations involved in explanatory depth (but are also related to evidential accuracy), their degree of integration is similar to my universal coherence, and finally their cognitive salience is similar to my internal coherence. Their work mostly illuminates explanatory depth. Their restricted scope—unpacking only explanatory power—is more limited than my systematization of all the major virtues. Furthermore, they do not identify a theoretical virtue that fails to fit within my systematization. They try to reduce two aesthetic virtues (simplicity and beauty) to non-aesthetic virtues that roughly fit within my coherent class of virtues (p. 212). This attempted reduction fails, as I show in Sect. 5, due to the close connections among beauty, simplicity, and unification (coupled with unification's widely acknowledged substantial epistemic status).

so far as possible, the number of types of facts we must accept as brute (1981, p. 529).

Explanations provide repeating, but contextually varied, motifs for understanding the world, unificationists argue. An aesthetic dimension lurks here. We will draw insight from this theory of explanation when we analyze the aesthetic theoretical virtues: beauty, simplicity, and unification. The other two major accounts of explanation, causal–mechanical and causal–counterfactual, are more widely held than the unificationist account (and recall that the two dominant theories of explanation provide some of the rationale for the evidential theoretical virtues of causal adequacy and explanatory depth). This fact, along with the *prima facie* epistemic priority of evidence in theory choice, contribute to the rationale for treating the evidential theoretical virtues first in my systematization. First place signifies epistemic priority for evidential virtues in theory formation and theory choice. Keep this in mind as we turn to our next class of virtues: those that are about how theoretical components fit together—cohere.

4 Coherential theoretical virtues: internal consistency, internal coherence, and universal coherence

4.1 Internal consistency (TV4)

Internal consistency is exhibited in a theory when its components are not contradictory. Douglas' (2013) taxonomy of theoretical virtues illuminates the significance of internal consistency and its relations to other virtues. So we begin with her account, and then correlate it with my systematization. Douglas organizes theory virtues into four groups based on whether they constitute minimal criteria or ideal traits and whether they pertain just to the theory *per se*, or to the theory in relation to evidence. Internal consistency, which is my first coherential virtue, belongs to group 1 of Douglas' taxonomy: minimal criteria applied only to the theory *per se*. At minimum a good theory must be internally consistent, especially in the sense of formal logical coherence. If a theory lacks this, then something about it is wrong (I qualify this below). Evidential accuracy, the first of the evidential virtues that we surveyed earlier, belongs to Douglas' group 2: minimal criteria applied to the theory in relation to evidence. But Douglas does not classify my other two evidential virtues—causal adequacy and explanatory depth. Perhaps this is because she aims primarily at articulating basic taxonomic principles illustrated with examples, rather than supplying a systematization of all the major theoretical virtues. I endeavor to do both.

Vickers (2013) argues that almost all purported examples of internal inconsistency in science fail to meet appropriate inconsistency criteria, chiefly (i) jointly employed propositions (ii) thought to be true (rather than approximations or idealizations) by a scientific community (iii) that involve a contradiction.⁵ He offers eight historical case studies to support the conclusion that there are many ways to avoid internal inconsistency as strictly defined. When internal inconsistency does occur, it is not

⁵ I thank a reviewer for pointing me to Vickers.

noticed until someone sees that a contradiction is derivable from the jointly employed propositions. Such contradiction recognition is difficult (as some case histories show) if the derivation is complicated or involves inference procedures that are unfamiliar or difficult to identify as truth-preserving.

Vickers and most of his critics agree that scientists highly value internal consistency, though what counts as compliance (see Vickers' three criteria above) is debated. Vickers' move to exempt idealized or approximated factors from significant inconsistency charges is particularly problematic, as much of science deals with approximations or idealizations. As [Frisch \(2016\)](#) put it in his review of Vickers:

When we have reasons to suspect that the inconsistency is the result of an idealization and to believe that the inconsistency would disappear if we de-idealized, the inconsistency may in fact be less important or interesting. But there are also cases that appear to involve idealizations or fictions for which we simply do not know whether the propositions in questions can be underwritten by a consistent de-idealized scheme. And these cases arguably present us with both scientific and philosophical puzzles.

Observations like this suggest that Vickers set the inconsistency bar too high, which predisposes us to underappreciate interesting inconsistency cases that would further show how competing theories might exhibit (or fail to exhibit to some degree) the theoretical virtue of internal consistency. Despite this weakness, Vickers' work greatly illuminates internal consistency in scientific practice.

Let us return to Douglas' work. Whereas in her taxonomy, internal consistency is the only member of her group 1, in my account it belongs to a group of coherential virtues that also include internal coherence and universal coherence. These three theoretical virtues concern how theoretical components fit together (cohere), but in three different ways. While internal consistency is restricted to logical coherence within the theory itself, the other two coherential virtues have progressively broader meaning. Let us explore the second and third coherential virtues in order to appreciate why they all belong to the same class of theoretical virtues, but also how they are each distinctive—and thus recognized appropriately as different virtues.

4.2 Internal coherence (TV5)

Internal coherence, the second coherential virtue, is possessed by a theory whose components are coordinated into an intuitively plausible whole. This virtue constitutes a kind of coherence that is more subtle and extensive than the logical principles of internal consistency. [Vickers \(2013, pp. 226–228\)](#) suggests that some purported cases of internal inconsistency might be plausibly identified as deficiencies in internal coherence: a set of propositions that were pointedly grouped by scientists in an *implausible* manner. In the heat of advocacy, scientists sometimes overstate the weaknesses of a rival theory, labelling it internally inconsistent (lacking TV4), while ignoring the possibility that it is only lacking some degree of internal coherence (TV5). [Schindler \(2014\)](#), in his analysis of Mendeleev's periodic table as an especially internally coherent classification theory of elements (despite its initial lack of evidential fit with a few

“known” elements), notes that internal coherence evades definition by necessary and sufficient conditions. However, it can be described using clear cases.

For clarity, consider the negative (vice) formulation of TV5: a theory lacks internal coherence to the extent that it incorporates ad hoc hypotheses. This typically refers to the construction of a theoretical component (hypothesis) that is attached to a theory in order to solve an isolated problem, but which is illegitimate in one or more other respects. Illegitimacy criteria include: it is insufficiently testable (e.g., too imprecise), it explains no other significant facts beyond the data that prompted its construction, and its “fit” within the larger theory is (to some degree) conceptually incoherent—awkward, arbitrary, or superficial.

Recent analyses of the status of ad hoc hypotheses range widely from articulating the relative clarity of ad hocness by means of surveying obvious cases of inadequate science or pseudoscience (Boudry 2013),⁶ to arguing that the designation “ad hoc” is hopelessly subjective. Hunt (2012) argues for the latter by reference to historical cases in which participants in specific theoretical disputes disagreed on what counts as ad hoc, and situations in which allegedly past clear cases of ad hocery were later reinterpreted to not be ad hoc due to subsequent discoveries. Note that cases of ad hoc accusation repeal based on new discoveries need not imply hopeless subjectivity about the designation of ad hocness. Epistemological refinements (e.g., better understanding of more theoretical virtues), coupled with new scientific discoveries, might reasonably settle certain debates that earlier were shrouded in unclear charges of ad hocery, such as the debate about whether earth is at rest in a cosmic center (see also the case against Hunt in Friederich 2014).

4.3 Universal coherence (TV6)

The third major coherential virtue, universal coherence, is present if a theory sits well with, or is not obviously contrary to, other warranted beliefs. If a theory is at odds with other well established knowledge, then all the worse for that theory. Incoherence in one’s belief system indicates something is wrong. A prominent example of a scientific theory that scored low in this virtue was the steady state theory, which rivaled big bang cosmology before the stunning success of the latter in the second half of the twentieth century. Steady state theory posited the continual creation of new matter in order to maintain constant average density throughout an allegedly eternal universe. Such an idea was commonly taken to be in conflict with the scientific understanding of the conservation of matter-energy—not to mention the strong metaphysical intuitions and arguments that stood behind such conservation laws in science.

The coherential theoretical virtues express how theoretical components fit together well, and they do so in a progressively expansive manner. The first coherential virtue, internal consistency, is about logical rigor. The second virtue in this class, internal coherence, concerns a broader sense of coherence: a theory whose components are coordinated into an intuitively plausible whole. Ad hoc hypotheses degrade this kind

⁶ Boudry notes: “Many critics of Freudian psychoanalysis have argued that the psychoanalytic habit of *ad hoc* reasoning stems from ... the malleable and incoherent conceptual structure of the theory.”

of coherential virtue. Finally, universal coherence refers to a theory that sits well within one's total knowledge, especially the knowledge most firmly justified and most comparable to the theory in question. The term “universal coherence” might be more apt than “external consistency” (Douglas 2013; McMullin 2014) because it better conveys a progression within this class of virtues toward more comprehensive coherence and it plays well with the way epistemologists speak. The term “conservatism” is inappropriate for this virtue for reasons Longino (1996) exposes. She also critiques distortions of this virtue that inhibit novel thinking.

The first three classes of theoretical virtues—evidential, coherential, and aesthetic—are arranged in decreasing order of epistemic weight (as typically judged). The position and character of the first two classes in this ranking is now clear. The aesthetic theoretical virtues might carry no *intrinsic* epistemic value. By this I mean that the aesthetic properties of theories might not (i) indicate the likely attainment of approximate truth or (ii) be a requirement for truth. Setting aside whether truth attainment in science is to be understood in a realist or antirealist manner, the evidential and coherential virtues are widely understood to be of intrinsic epistemic value—each of them is either a truth requirement or indicates the likely attainment of approximate truth (likewise for the diachronic virtues when, for example, they involve predictions that later are shown to have been approximately true beliefs about the future). We will explore whether the aesthetic virtues of theories have any epistemic value, and if so, whether this value is intrinsic or *extrinsic* (i.e., promotes, without indicating, truth attainment).⁷

5 Aesthetic theoretical virtues: beauty, simplicity, and unification

The aesthetic theoretical virtues possess an aesthetic shape (*fittingness*) that is qualitatively different from the logical-conceptual *fit* of the coherential virtues. Scholars in many fields sometimes appeal to the aesthetic properties of theories in their case for accepting such theories. However, the epistemic status of the aesthetic virtues has been challenged more than that of any other virtues. Besides beauty and simplicity, which are the most frequently purported aesthetic theoretical virtues (Mackonis 2013, p. 979), I will argue that a certain sense of unification also belongs to this aesthetic class of virtues. Unification has more widely recognized *intrinsic* epistemic value (recall Kitcher's unificationist account of explanation in Sect. 3.4) than either beauty or simplicity. We will also explore the possibility of finding an epistemic role for beauty by arguing descriptively from the perceptions of some prominent scientists and from the contention that simplicity and unification are epistemically enhanced special cases of beauty. Lipton (2004, p. 68) came close to recognizing this aesthetic class of theoretical virtues when he wrote in regard to his theory of inference:

⁷ I adapted Steel (2010) who argues that a theoretical virtue “is intrinsically epistemic if exemplifying that value either constitutes an attainment of truth or is a necessary condition for a statement to be true.” Such “truth does not necessarily mean true theories. For instance, the ability to make accurate predictions is an example of the attainment of truth, since accurate predictions are true or approximately true statements about the future.” He contends, using simplicity as an example, that a theoretical virtue has *extrinsic* epistemic value if it promotes truth attainment without itself being an indicator or requirement of truth.

Moreover, if we do end up selecting Inference to the Best Explanation, it will not simply be because it seems the likeliest explanation, but because it has the features of unification, elegance, and simplicity that make it the loveliest explanation of our inductive behavior.

5.1 Beauty as the most general aesthetic theoretical virtue (TV7)

According to my account of aesthetic theoretical virtues, a beautiful theory evokes aesthetic pleasure in properly functioning and sufficiently informed persons (with some degree of cultural and individual variation of aesthetic experience). The properties of theories and mathematical proofs that are among those factors that trigger the experience of beauty include symmetry, aptness (McAllister 1996, pp. 172–173) and surprising inevitability (Montano 2014, pp. 34–36).

In contrast, aesthetic relativism suggests that no judgments about beauty or ugliness (whether in regard to a theory or anything else) are more correct than others. If this view of aesthetics were correct, then it is difficult to see how any aesthetic theoretical virtues could be of rational importance in the theory choice. However, there are good reasons to reject aesthetic relativism. We often make aesthetic judgments and take them to be at least approximately correct, especially as we mature as persons. So aesthetic relativism is out of step with common practice. Moreover, Zangwill (2014) notes that

... one can virtually always catch the professed relativist about judgments of beauty making and acting on non-relative judgments of beauty—for example, in their judgments about music, nature and everyday household objects. Relativists do not practice what they preach.

Zangwill (2014) pinpoints another disturbing feature of aesthetic relativism.

For if “it’s all relative” and no judgment is better than any other, then relativists put *their* judgments wholly beyond criticism, and they cannot err. Only those who think that there is a right and wrong in judgment can modestly admit that they might be wrong. What looks like an ideology of tolerance is, in fact, the very opposite. Thus relativism is hypocritical and it is intolerant.

Due to a lack of space, we will explore only a few of the many perspectives on aesthetics that are relevant to theoretical virtues. Other views that we will not address include Breitenbach’s (2013) Kantian theory, McAllister’s (1996) largely projectivist account, and Montano’s (2013) major revision of McAllister. Subsequent work could explore how such various understandings of aesthetics might bear upon my taxonomy of theoretical virtues. This endeavor might include better grasping the demarcation of aesthetic properties from non-aesthetic properties. But, as Levinson (2003, p. 12) observes: “It is widely agreed that aesthetic properties are perceptual properties, dependent on lower level perceptual properties, directly experienced rather than inferred, and linked in some way to the aesthetic value of the objects possessing them.” He also identifies substantial agreement on clear examples of aesthetic properties. Note how he begins his “open-ended” list with “beauty” (and its opposite ugliness), and

notice also the lurking presence of both simplicity and unification highlighted by my italics.

... beauty, ugliness, sublimity, grace [*simplicity* or refinement], elegance, delicacy, harmony, balance, *unity*, power, drive, élan, ebullience, wittiness, vehemence, garishness, gaudiness, acerbity, anguish, sadness, tranquility, cheerfulness, crudity, serenity, wiriness, comicality, flamboyance, languor, melancholy, sentimentality—bearing in mind, of course, that many of the properties on this list are aesthetic properties only when the terms designating them are understood figuratively (p. 6).

Benovsky (2013) argues that all aesthetic evaluative properties of theories (his terminology for aesthetic theoretical virtues) are grounded in the non-aesthetic evaluative properties of theories such as internal consistency, internal coherence (intuitive plausibility), universal coherence, explanatory power, and simplicity. The first three non-aesthetic evaluative properties listed here are the three coherentist virtues in the previous section of my essay. Explanatory power is an ambiguous term that often refers to various non-predictive theoretical virtues, but that most often refers to the evidential virtues, especially causal adequacy and explanatory depth—the two evidential virtues that go beyond mere evidential fit. Simplicity is difficult to characterize. Many, including myself for reasons I will articulate, classify it primarily as an aesthetic property of theories, rather than as a non-aesthetic evaluative property as does Benovsky. However, according to my account, persons make *partially aesthetic judgments* when they differentially weight, or aptly balance, all of the theoretical virtues in particular episodes of theory choice and theory refinement. Indeed, aptness and balance are aesthetic properties. Non-aesthetic considerations also figure into the differential weighting of the virtues in theory choice, but this is beyond the scope of the present study.

Scientists, mathematicians, philosophers, and other scholars sometimes appeal to their aesthetic engagement with theories—experienced most generally as an encounter with beauty—in order to help justify theory choice. Although this descriptive argument for the virtuousness of theoretical beauty is weak on its own, it is worth consideration because later I will argue that simplicity and unification are special cases of beauty that more likely carry at least some epistemic weight. Dirac, one of the founders of quantum mechanics, celebrated general relativity's elegance.

The foundations of the theory are, I believe, stronger than what one could get simply from the support of experimental evidence. The real foundations come from the great beauty of the theory.... It is the essential beauty of the theory which I feel is the real reason for believing in it (1980, p. 10).

McCartney expresses the affinities of the three major aesthetic theoretical virtues as seen in physics.

Anyone who has dealt with Maxwell's equations of electromagnetism, or the Schrödinger equation, cannot fail to be impressed with the concise elegance of the formulae. Their compact beauty is made all the more stark when one considers

the very broad range of phenomena they ultimately explain and encapsulate (2015, p. 3).

Beauty and (as I shall argue) its two most epistemically important special cases of simplicity and unification are vividly displayed here in their mutual relations. Educated contemplation of the mathematical formulae of physics occasions an encounter with beauty—an experience that is partially describable in terms of both the compactness (simplicity) of the formulae, and the remarkable range of natural phenomena they explain (unification).

Although generic theoretical beauty might have no intrinsic epistemic value, it more likely possesses extrinsic epistemic value to the extent that such a general aesthetic experience inclines researchers toward recognizing and cultivating simplicity and unification as special kinds of beauty that are more epistemically relevant in theory choice. Mackonis (2013, p. 979) remarks regarding “the meaning of beauty or elegance” in theory choice: “simplicity ... is both one of the most often cited explanatory virtues and one of the most often cited features of beauty.” Let us explore this further.

5.2 Simplicity (TV8) and unification (TV9) as specific complementary aesthetic theoretical virtues

Now that we have an introductory grasp of beauty in theories as a general aesthetic virtue, we are ready to characterize the specific aesthetic virtues of simplicity and unification as special cases of beauty that are particularly important for theory choice and theory refinement. A theory that exhibits simplicity explains the *same facts* as rival theories, but with *less* theoretical content. A unified theory, however, is one that explains *more kinds of facts* than rival theories with the *same* amount of theoretical content (Thagard 1978). Simplicity and unification address the same thing, the style of informativeness, from opposite complementary orientations. Simplicity is increased informativeness by means of a comparative *reduction* (relative to rival theories) of theoretical content. Unification is increased informativeness by means of a comparative *increase* in the different kinds of data that get explained. A theory can be evaluated (compared to rivals) for its informativeness in proportion to its theoretical content in both stylistic directions—simplicity and unification.⁸ Such comparative evaluations may be difficult to characterize formally due to their subtlety. However, in the case of simplicity, “less theoretical content” means, roughly, fewer entities postulated by the theory (often called parsimony or ontological simplicity) and/or fewer or more concise basic theoretical principles (often called elegance or syntactic simplicity). Note the aesthetic term “elegance,” which reflects simplicity’s aesthetic character.

In what sense are the complementary virtues of simplicity and unification (possibly) intrinsically *epistemic* in an *aesthetic* manner? Lipton (2004, p. 124) remarked in connection with our inferential preference for simple and unified explanations: “if the

⁸ Mackonis (2013, p. 987) develops a similar account of simplicity and unification: “Simplicity is similar to unification in the sense that a hypothesis is better the less it takes to explain something. Unification asks to explain more facts with [the] same resources. Simplicity asks to explain [the] same facts with fewer resources.”

world is a chaotic, disunified place, then I would say it is less comprehensible than if it is simple and unified. Some possible worlds make more sense than others.” The comprehensibility of the world (which supports the epistemic aims of intelligent beings living in such a world) has multiple aesthetic dimensions, including most prominently, simplicity and unification. We prefer to live in a comprehensible world because this is more pleasing to the mind. The history of science attests to this aesthetic preoccupation and its respectable track record in scientific discovery (Glynn 2010). In such a highly comprehensible world, theories ranking higher in simplicity and unification are more likely to be approximately true than rivals that rank lower (the opposite would be true in many more other possible worlds). There is nothing necessary about the existence of a world characterized by simplicity and unification, but our minds are preoccupied by the desire to find such a simple unified world. This aesthetic orientation operates in the broader culture. Simplicity and unification are among the aesthetic qualities celebrated in literature, film, and many other cultural arenas. Humans highly value a story that compactly instantiates a plot (simplicity) while expressing the voice of human experience in a broadly significant manner (unification). Moreover, historians craft stories that likely resemble “what happened” to the degree that they possess many of the theory virtues.

Many versions of simplicity have been identified and they include factors such as a theory that: postulates fewer entities, postulates fewer kinds of entities, raises fewer additional explanatory questions, and posits fewer primitive explanatory ideas (Beebe 2009). Swinburne (1997) adds: postulates fewer laws and postulates laws relating fewer variables. McAllister (1996) observes that this multiplicity of simplicity criteria—many of which might be relevant to a particular theoretical dispute—defies complete reduction to quantified evaluative procedures such as those Sober (2015) and Kelly (2011, 2016) propose.

Sober (2015) recommends simplicity as a consideration in theory choice *if* it is understood as the number of adjustable model parameters as treated within either the Bayesian or frequentist philosophical traditions of probability theory. He refers to such narrowed conceptions of simplicity as “parsimony” and he illustrates their epistemic value (and instances where they do not apply) in case studies spanning evolutionary biology, psychology, and philosophy. “I am a reductionist about parsimony,” he announces. In short, he considers the aesthetic value of simplicity to be merely subjective and irrelevant to the epistemic value of simplicity.

Sober aims to show that simplicity helps scholars to infer which explanation is best, even if reality is not structured by beautiful simple principles. Several (but not all, says Sober) of the mathematically rigorous treatments of simplicity “agree that parsimony, as measured by the number of adjustable parameters in a model, is relevant to making ... estimates” of either a model’s predictive accuracy or its likelihood (p. 141). In either case simplicity “is not a subjective aesthetic frill. It has an objective epistemic status” (p. 147). While tentatively applauding with Sober these achievements of probability theory, we should remember, as Sober tacitly concedes, that there might also be *epistemic-by-means-of-aesthetic* value in theoretical simplicity that is not addressed by mathematical analysis. But due to Sober’s reductionist agenda, he dismisses as “subjective aesthetic frill” (sounding like an aesthetic relativist) the purportedly larger

aesthetic dimension of rational theory choice (in simplicity and beyond) that has been proclaimed by many influential scientists, including Einstein and Dirac.

Steel (2010, p. 19) notes that, if Sober or Kelly is correct, then simplicity is an extrinsic epistemic value “even if it is not an intrinsic one” (supposing like Sober that its intrinsic epistemic value would require that the world itself is dominated by simplicity; note how DeLancey disputes this assumption below). A theoretical virtue has extrinsic epistemic value, Steel argues, if it promotes the attainment of truth without itself being an indicator or requirement of truth. We have already seen how Sober made such a case for the extrinsic epistemic value of simplicity. Steel summarizes Kelly’s (2011, 2016) alternative to Sober:

A second account of simplicity explains how a preference for simpler hypotheses promotes efficient convergence to the truth, where efficiency is understood in terms of minimizing the maximum number of times the investigator can switch from conjecturing one hypothesis to another.... Both of these accounts, then, defend simplicity as an extrinsic epistemic value. In both cases, a preference for simpler hypotheses is argued to promote the attainment of truth (either of approximately true predictions or efficient convergence to true hypotheses), yet neither approach presumes that the world is simple.

Forster and Sober (1994, p. 14) treat unification’s epistemic role in theory choice in a manner that is similar to their probabilistic analysis of simplicity.

We conclude that estimated accuracy explains why a unified model is (sometimes) preferable to its disunified competitors. At least for cases that can be analyzed in the way just described, it is gratuitous to invoke unification as a *sui generis* constraint on theorizing.

Forster and Sober concede here that many instances of unification are *not* captured by mathematical model selection procedures that might have reduced the rationality of favoring more unified theories to merely their higher predictive accuracy estimates. So Forster and Sober concede that not all cases of simplicity and unification are reducible to non-aesthetic predictive accuracy or likelihood. They leave open the possibility that there are irreducibly aesthetic aspects to simplicity and unification that rationally allure us to theories that possess them. Indeed, many scientists have believed that aesthetic properties have a legitimate modest role in rational theory choice, and such belief has guided successful scientific practice in some cases (Glynn 2010). Although the aesthetic branch of my taxonomic argument does not resolve this issue, let us now explore additional pointers that might help guide work in this area.

DeLancey (2011) argues that the simplicity criterion in theory choice could operate reasonably whether or not the world is elegantly simple (and he does so in a manner that differs from Sober and Kelly). His case comes with an ontological stipulation that he thinks is inoffensive to most antirealists: “a minimal realism about the existence of objects and laws, in order to allow that the descriptions of the relevant phenomena contain patterns.” After arguing that we have no reason to expect that the world’s potential complexity has any definite upper limit, DeLancey maintains that “simple” is a relative term, like “small.” So, for any magnitude of complexity “it’s reasonable to call it ‘simple,’ since there are infinitely many alternatives that are more complex”

(p. 94). This means that simplicity only operates comparatively with respect to rival theories, which is how I have characterized it. Huemer (2009, p. 219) offers a similar account of simplicity (alongside three other accounts).

The boundary asymmetry account starts from the observation that there is a lower bound but no upper bound to the degree of complexity a theory can have. That is, for any given phenomenon, there is a simplest theory (allowing ties for simplest), but no most complex theory of the phenomenon: however complex a theory is, it is always possible to devise a more complicated one. This is most easily seen if we take a theory's complexity to be measured by the number of entities that it posits: one cannot posit fewer than zero entities, but for any number n , one could posit more than n entities. Similar points hold for other measures of complexity, such as the number of parameters in an equation.

The boundary asymmetry analysis of simplicity—and its stated rarity of theoretical simplicity compared to theoretical complexity—is supportive of simplicity's epistemic value (in contrast to the epistemically vicious greater complexity of indefinitely numerous rival theories). This contention stems, in part, from the observation that rarity is a common trait of epistemic and aesthetic value. The analytic project of identifying the necessary and sufficient conditions for a belief to count as knowledge has exposed countless ways for beliefs to fall short of knowledge, and comparatively few ways for beliefs to constitute knowledge. Rarity also is a prominent trait of aesthetic value. Beauty requires a kind of coordination (among musical notes, words on a page, or brush strokes on a canvass) that is rare compared to the vast number of possibilities for disharmony. In sum, DeLancey's and Huemer's accounts of theoretical simplicity (both of which are independent of how elegantly simple the world may or may not be) are compatible with the possibility that the value of simplicity in theories consists of interlocked aesthetic-epistemic properties.

Simplicity, although primarily an aesthetic theoretical virtue, has secondary affinities to other virtue classes—we will cover the coherential affinity now and the evidential one in Sect. 7. When we explored internal coherence (TV5), we noted that a theory lacks internal coherence to the extent that it incorporates ad hoc hypotheses that are illegitimate in one or more respects, including insufficient precision and insufficient conceptual fit within the larger theory—often resulting in an awkwardly complex theoretical monstrosity. Note how internal coherence (lack of ad hocness) is similar to simplicity: one way for a theory to score low in simplicity is by the presence of ad hoc hypotheses. So simplicity overlaps with internal coherence. Of course, simplicity also decreases by the addition of non ad hoc hypotheses. So internal coherence and simplicity are quite different.

My account of the aesthetic virtues sits well within Zangwill's aesthetic realism (2001, 2014). He argues that specific substantive aesthetic properties such as delicacy and dumpiness are *ways of being* beautiful or ugly—which he calls verdictive aesthetic properties. The latter refer to the degrees of overarching aesthetic merit (beauty) or demerit (ugliness). Adapting Zangwill's metaphysics to the present study, the theoretical virtue of beauty is the aesthetic property that designates the overall aesthetic value of a theory as grounded in specific substantive aesthetic properties such as symmetry and aptness—which are some of the ways of being theoretically beautiful that may

lack the epistemic potency of simplicity and unification. However, simplicity and unification are two substantive aesthetic properties of theories that stand out from the rest due to their interlocking complementarity and (according to many scientists) greater epistemic credentials compared to other substantive aesthetic properties of theories (e.g., symmetry and aptness).

Zangwill does not think that theoretical “beauty” in mathematics and science refers to actual aesthetic judgments, but this is a minor error in an otherwise insightful book (2001, pp. 140–142). He argues that when a scientist describes a theory as beautiful, that such language is metaphorical, and only means that the theory achieves the aim of explaining the data well. But there are counterexamples—beautiful hypotheses that did *not* explain well the relevant data. For example, Kepler toyed with multiple possible beautiful closed curved figures (geometrical hypotheses) to see which would best fit the astronomical data of planetary positions (he focused on Mars). He finally settled on the ellipse, which exhibits less simplicity than the traditional circular heavenly motions that even Copernicus had retained. Of course Kepler’s new astronomy exhibited heightened simplicity in other respects. Indeed, Kepler often appealed explicitly to the aesthetic properties of theories as partial grounds for their likely truth, a practice that made sense to him because God “introduced nothing into Nature without thoroughly foreseeing not only its necessity but its beauty and power to delight” (Kepler 1981, p. 55). Many other scientists and philosophers, holding a diversity of religious and non-religious views, also have treated beauty in scientific theories as truly aesthetic *and* epistemic (Breitenbach 2013).

Let us summarize and reinforce the justification for classifying unification as an aesthetic theoretical virtue alongside simplicity. As I have argued above, simplicity and unification address the same thing, style of informativeness, from opposite complementary orientations. Given that simplicity is the most often cited example of an aesthetic theoretical virtue, and given the widely recognized complementary affinity between simplicity and unification, it is plausible to consider unification to be an aesthetic theoretical virtue that assists (in a complementary fashion with simplicity) in rational theory choice. Given that unification is the centerpiece of one of the leading theories of explanation (as outlined earlier), what does this imply regarding unification constituting an aesthetic theoretical virtue?

A unificationist account of explanation would have us give ultimate priority to “using a few patterns of argument in the derivation of many beliefs” because we thereby “minimize the number of types of premises we must take as underived,” which reduces “the number of types of facts we must accept as brute” Kitcher (1981, p. 529) According to this account of explanation, unification is equated with explanation itself, rather than recognizing unification as an aesthetic component to what helps make an explanation more likely true. But if one attends carefully to Kitcher’s description of his own unificationist theory of explanation, some of his descriptive language has an aesthetic ring to it. Furthermore, the different strengths of each of the major theories of explanation (unificationist, causal–mechanical, and causal–counterfactual) might urge us to resist the temptation to reduce all explanation to any one of these explanatory approaches.

Because some scholars are attracted to a unificationist account of explanation (despite its greater weaknesses compared to the other two major accounts), this

helps us to recognize unification as a theory virtue that might have some modest epistemic credentials. And given unification's interlocking complementarity with simplicity (and the arguments for the quantitative correlation between *certain facets* of simplicity/unification and estimates of predictive accuracy, likelihood, or efficient convergence on truth), this offers some support for the epistemic standing of simplicity. Furthermore, the aesthetic theoretical virtue of beauty might be shown to have more than zero epistemic value if, as I have argued, unification and simplicity are the two main epistemically valuable ways of being beautiful. Even so, in general, the aesthetic virtues are widely considered to have less epistemic value (if any) than either of the first two classes of theoretical virtues: evidential and coherential. If my argument for the possible epistemic role of the aesthetic theoretical virtues lacks sufficient force to move certain readers, I would offer them this aesthetic taxonomic class as merely descriptive of some scientific practice (Glynn 2010)—and I would note how in Sect. 6.2.1 the diachronic sense of unification constitutes a mode of fruitfulness. Finally, even if some readers find my description of these virtues unhelpful, I would still recommend to them the rest of my taxonomy, which stands on its own even in the absence of the aesthetic class.

The fourth class of theoretical virtues possesses a distinctive temporal dimension that is missing in the three previous classes—evidential, coherential, and aesthetic. The diachronic virtues come last in my systematization because they cannot be instantiated in the initial framing of a theory, as is true of the theoretical virtues in the other three classes. Diachronic virtues require additional time *after* a theory is launched. Their time has come.

6 Diachronic theoretical virtues: durability, fruitfulness, and applicability

Durability, fruitfulness, and applicability, which I recognize as the chief diachronic theoretical virtues, can only be instantiated as a theory is cultivated *after* its origin. This necessarily extended temporal dimension of the diachronic virtues is, arguably, of considerable epistemic importance. But even if one endorses the arguments that discount the epistemic significance of this temporal component (Mayo 2014), one still should acknowledge a group of virtues that (unlike the other theoretical virtues) can only be instantiated in a theory *after* its initial formulation. Time is of their essence in a manner that goes beyond the trivial truth that all human endeavor is temporal. Keep this defining feature of the diachronic virtue class in mind as we survey its principal members, otherwise these virtues might appear arbitrarily tossed into an “other” category plagued by wild heterogeneity. McMullin (2014) has led the way in articulating the epistemic significance of two of the three main diachronic virtues: durability and fruitfulness (I recognize McMullin's third diachronic virtue of “consilience” as a mode of fruitfulness). Axtell (2014, p. 232) observed that “the most unusual and useful feature of McMullin's taxonomy is his close attention to ‘Diachronic’ theory virtues.” Applicability, largely overlooked as a theory virtue, is another important member of this diachronic category, as I shall demonstrate.

6.1 Durability (TV10)

A theory exhibits durability if it has survived testing by successful prediction or by plausible accommodation of new unanticipated data (or both). Popular or long-lived theories are not necessarily durable in the epistemic sense in view here. Equating durability with popularity or tradition is fallacious. Testability is a prerequisite for (and a potential constituent of) durability, but many testable theories have failed too many tests to be durable. The more testable a theory is, the more durable it would prove itself to be if it passes the tests.

Despite the important role of predictive success as a component of durability in many areas of science, it is less prominent in some reputable scientific theories that are, nevertheless, well endowed with other virtues. Successful prediction is very frequently part of explaining “how things work,” but somewhat less routine in explaining “how things originated”—as in theories about the history of the cosmos, earth, and life (Cleland 2011, but Williams 1973 and Winther 2009 argue otherwise). Successful historical theories very frequently show their durability by a track record of plausible accommodation of new data that, although not predicted, came to light after the theory’s origin. The durability of a theory suffers if one or more of its predictions are disconfirmed or when theorists respond to disconfirming evidence by modifying the theory with ad hoc hypotheses (see Sect 4.2). Although initially a theory may exhibit evidential accuracy and many other non-diachronic virtues, it is impossible for a newborn theory to instantiate the virtue of durability—this *takes time* in a sense not required by the non-diachronic virtues. A similar necessary temporal dimension characterizes fruitfulness. We will learn more about durability as we compare it with fruitfulness.

6.2 Fruitfulness (TV11)

Fruitfulness, also known as fertility or fecundity,⁹ is another diachronic theoretical virtue. A theory is fruitful if, over time, it generates additional discovery by means such as successful novel prediction, unification, and non ad hoc theoretical elaboration. While durability is about conservation (a theory passing tests to survive), fruitfulness is about innovation (a theory stimulating further discovery). When a prediction formulated in the context of a theory’s construction is later verified, this successful predictive outcome increases the virtue of durability in that theory. By contrast, a *novel* prediction is one that was not conceived in conjunction with a theory’s construction, but that nevertheless follows reasonably from it. When such a novel prediction is confirmed by observation, a theory exhibits fruitfulness.¹⁰

⁹ The latter two terms might convey a theory’s *likely potential* to generate new discovery, rather than a proven track record of actually doing so. Thus “fruitfulness” is a more appropriate term to refer to this kind of diachronic theoretical virtue.

¹⁰ McMullin (2014, p. 505) defines a novel prediction as “one whose success would count as unexpected were the postulated structure indeed to be a fiction, unexpected because the novel result lies to some degree outside the scope of the data originally accommodated by the theory. The farther outside, the more unexpected it would be and the stronger the confirmation it would offer for the theory’s ontological grounding.” Compare

The closely related diachronic character of durability and fruitfulness is well illustrated in the discovery of the first two planets beyond Saturn. Soon after Friedrich William Herschel unexpectedly discovered Uranus in 1781, astronomers noted that its observed motion strayed from what contemporary Newtonian mechanics predicted of such a planet. However, given the overall theoretically virtuous status of Newtonian physics up through that time (including its durability due to its success in testing), most astronomers expected a forthcoming way to make Uranus compliant with established theory. Even rejecting the anomalous data as “inaccurate” seemed reasonable early on. By the 1830s, however, the possibility of a perturbing planet beyond Uranus became a more reasonable and popular speculation, despite the absence of a precise novel prediction of where to find such a planet. By this time many astronomers were mostly confident in the accumulated data of Uranus’ positions in the sky.

This brings us to the celebrated successful novel prediction of 1845–1846. Based principally on Newtonian physics and the well-known irregularities in Uranus’ motion, two astronomers independently predicted where another unknown perturbing planet (later called Neptune) was likely located. Le Verrier’s estimate of the planet’s location was the most accurate (correct within one degree), as confirmed by a German astronomer in 1846. The (*fruitful*) novel prediction of Neptune was born within the context of a *durable* Newtonian orbital mechanics research tradition and the unexpected discovery of Uranus with its anomalous motions. The sensational success of this novel prediction (the discovery of Neptune) also rendered Uranus a Newtonian-compliant planet—thus further vindicating earlier provisional toleration of Uranus’ anomalies, a toleration that had been justified by yet earlier Newtonian durability and fruitfulness.

Smith’s (2010, 2014) study of gravity theory from Newton to the present further illuminates the durability and fruitfulness of this research tradition, and it includes the case histories of Uranus and Neptune. Smith was surprised that the principal kind of question being tested was not “Do the calculated motions [e.g., of Uranus] agree with the observed motions?” (which is a question of durability). Rather it was: “Can robust physical sources compatible with Newtonian theory be found for each clear, systematic discrepancy between the calculated and the observed motions?” (which is a question of fruitfulness). Neptune (as novelly predicted) turned out to be such a robust physical source. However scientists failed over a half century to find a robust (detectable) physical source for the Newtonian-defying behavior of Mercury—a tiny anomaly in the precession of its perihelion. But this failure, which Einstein solved by way of theory replacement, might not completely diminish the epistemic significance of two centuries of Newtonian durability and fruitfulness. Smith notes: “All the other discrepancies ended up revealing some detail of our planetary system, the least subtle of which was Neptune, that theretofore had not been taken into account in the calculations” (2010, p. 552).

Such serial Newtonian problem solving shows an interlocking of durability (passing tests to survive) and fruitfulness (stimulating further discovery). For example, Uranus’

Footnote 10 continued

Nolan (1999) who denies fruitfulness as a distinct theoretical virtue that is supportive of McMullin’s case for scientific realism, and the realist response by Segall (2008).

temporarily Newtonian-defying behavior “would have been masked if the significantly larger gravitational effects of Saturn on Uranus had not been included in the calculation first.” Smith explains further:

So, the discovery of Neptune provided evidence not only for Newton’s theory, but also for the specific aspects of Saturn that entered into calculating its effects on Uranus, for these were no less presupposed in the anomaly that emerged than Newton’s theory was. The point generalizes. Each time a discrepancy emerges and a robust physical source for it is found, that source is incorporated into the new calculations, and the process is repeated, typically with still smaller discrepancies emerging that were often theretofore masked in the calculations. So, what was being tested each time when a new discrepancy emerged and a physical source for it was being sought was not only Newtonian theory, but also all the previously identified details that make a difference and the differences they were said to make without which the further systematic discrepancy would not have emerged (2010, pp. 552–553).

Though some philosophers have argued to the contrary (Collins 1994; Harker 2008), many scientists and philosophers think that predictive success—especially novel predictive success—is a stronger indicator of likely approximate truth than a theory’s accommodation of data (Douglas and Magnus 2013). According to my systematization (which illuminates but does not settle this thorny issue), data accommodation refers to a theory’s initial instantiation of the evidential virtues (evidential accuracy, causal adequacy, and explanatory depth), and a theory’s subsequent instantiation of certain diachronic virtues, namely non-predictive durability (plausibly making sense of new unanticipated data) and non-predictive fruitfulness (especially non ad hoc theoretical elaboration that makes sense of new unanticipated data).

6.2.1 Unification as a mode of fruitfulness

Fruitful theory elaboration, whether by means of successful novel prediction or non ad hoc theoretical elaboration that makes sense of unanticipated evidence, often also makes sense of *new kinds* of data, and thus is additionally recognized as increasing a theory’s unification. Earlier we encountered unification as a non-diachronic (aesthetic) theoretical virtue. The diachronic increase of unification differs somewhat from its non-diachronic cousin. The historian and philosopher of science William Whewell (1794–1866) called diachronic unification “consilience.” When a theory explains a new domain of facts in a surprising way, then it is fruitful in a consilient manner. McMullin writes in this regard:

A good theory will often display remarkable powers of unification, making different classes of phenomena “leap together” over the course of time. Domains previously thought to be disparate now become one, the textbook example, of course, being Maxwell’s unification of magnetism, electricity, and light. Examples abound in recent science, a particularly striking one being the development of the plate-tectonic model in geology. Assuming that this unifying power manifests itself over time, it testifies to the epistemic resources of the original theory

and hence to that theory's having been more than mere accommodation (2014, p. 505).

McMullin contrasts diachronic unification with its non-diachronic counterpart: "If the unification was achieved by the original theory, however, the virtue involved would no longer be diachronic." Instead, it would count (in my systematization) as an aesthetic theoretical virtue that I simply call "unification," and that Lipton calls "variety" (and yet others call "broad scope"). Lipton favors the assumption that such "heterogeneous evidence provides more support than the same amount of very similar evidence" (Lipton 2004, p. 168). Despite my own inclination to accept Lipton's point, I recognize this as a somewhat debatable assumption about the epistemic significance of an aesthetic property. However, when unification increases over time, especially by means of surprising convergences, then unification is less likely the result of the idiosyncratic aesthetic predispositions and clever accommodating skills of a theorist during theory formation. Thus fruitful diachronic unification has greater epistemic value than a theory's initial degree of aesthetic unification. In this regard my theoretical virtue systematization makes better sense of McMullin's argument than McMullin's own systematization in which aesthetic theoretical virtues are not recognized.

6.2.2 *The role of prediction in the diachronic virtues*

Drawing from Douglas' work on the relationship of prediction to inferring the best explanation, I argue that predictive success (in the first two diachronic virtues explored above) extends the epistemic work of many non-diachronic theoretical virtues such as causal adequacy, explanatory depth, beauty, simplicity, and unification. These latter theory traits, which she collectively labels as "explanatory,"

...appeal to us, not just because we are aesthetically driven creatures but because such virtues help us to use the explanation to think and, in particular, to think our way through to new predictions, new tests, new rigors for our beautiful explanation (2009, p. 460).

Douglas also notes:

Predictions are valuable because they force us (when followed through) to test our theories, because they have the potential to expand our knowledge into new realms and because they hold out the possibility (if successful) of gaining some measure of control over natural processes (2009, p. 455).

Transposing Douglas' insights into my taxonomic terms, predictions are valuable because they figure into all three of the major diachronic virtues: durability (testing theories successfully), fruitfulness (expanding "our knowledge into new realms"), and applicability (which includes "gaining some measure of control over natural processes"). Moreover, the operation of prediction ("saying before" at least in a logical if not temporal sense) in these three theoretical virtues further supports my classification of them as diachronic. We have surveyed the first two members of the diachronic class (durability and fruitfulness), and are now ready to investigate the final diachronic

virtue of applicability. This is the last major theoretical virtue addressed in my systematization.

6.3 Applicability (TV12)

Applicability refers to when a theory is used to guide successful action (e.g., prepare for a natural disaster) or to enhance technological control (e.g., genetic engineering). High degrees of the virtue of applicability obtain when a theory that is used to guide such action or control provides more effective outcomes than what is possible in the absence of the theory. Successful scientific theories constitute *knowledge* of the world (knowing *that*), not *control* over the world (which is mainly knowing *how*) for practical (non-theoretical) purposes.¹¹ In this regard [Strevens \(2008, p. 3\)](#) notes: “If science provides anything of intrinsic value, it is explanation. Prediction and control are useful ... but when science is pursued as an end rather than as a means, it is for the sake of understanding.” But even after the intrinsic good of a theoretically virtuous explanation is in hand, one of several possible additional confirmatory diachronic (predictive or controlling) virtues might be acquired by a theory, including applicability. In such cases a good theory just gets better.

Although scientific experiments use technological control, they do so to test scientific theories—so the main function is still to understand nature, not to control it. However, especially in the case of theories supported by experimentally verified prediction, such foreknowledge and laboratory control might be exploited to achieve practical aims such as device fabrication or medical intervention. But in any case, one cannot *apply* scientific knowledge until *after* one first *obtains* it. This necessary time lapse makes applicability diachronic.

To obtain scientific knowledge we search for a theory that (initially) exhibits many of the non-diachronic theoretical virtues. Subsequent work aimed at theory testing and elaboration might produce the additionally confirming presence of the diachronic virtues of durability and fruitfulness. At some point in this dance of virtue-driven theory assessment and refinement, sufficient confidence in a particular theory might spur attempts to apply it as the basis for a new or improved technology. If the derived science-based technology actually works, then the “applied theory” has acquired the additional theoretical virtue of applicability. Because this requires additional time after initial theory formation, the diachronic classification of applicability is appropriate.

Although the application of scientific theories constitutes one aspect of technology, much of technology involves the empirical discovery of “know how” knowledge without crucially presupposing or immediately applying any particular scientific theory. In his landmark study, [Vincenti \(1990\)](#) explains how even engineers, as the most sophisticated technologists, often acquire “know how” knowledge without applying scientific theory (theoretical “knowing that”). He analyzes several case studies in early aeronautical engineering to show that to a large degree this technological disci-

¹¹ [Vincenti \(1990, p. 237\)](#) argues that although engineering is primarily about procedural knowledge (knowing how) “it also requires a great deal of descriptive knowledge, which is synonymous with ‘knowing that’ or knowledge of how things are. Part of this knowledge comes from science, but much of it ... arises within engineering itself.”

pline was independent from the physical sciences. Although some of this aeronautical “know how” knowledge (e.g., improvements in wing airfoils and propeller shapes) was acquired by sheer trial and error, often they were constructed systematically using sophisticated mathematical models created by engineers without applying the natural sciences. However, [Vincenti \(1990, p. 230\)](#) acknowledges that “while engineering design is an art, it is an art that utilizes (increasingly) knowledge from developed and developing science.” But, he is quick to qualify his admission: “This is a far thing from saying, however, that science is the sole (or even major) source and that engineering is essentially applied science.” Indeed, the relation between science and technology is not a simple one-way linear affair ([Radder 2009](#); [Douglas 2014](#)). But this “emancipation” of technology from subordination to science, accomplished by historians and philosophers of technology between 1960 and 1990 ([Houkes 2009, p. 310](#)), should not obscure the epistemic significance of instances of technological innovation made possible, in part, by *applied* scientific theory.

This point is in harmony with the so-called demise of the “pure versus applied science” dichotomy. Understanding and controlling nature are closely related, as our study of the diachronic theoretical virtues, including applicability, indicates. Douglas (2014, p. 62) displays some of the subtlety of this argument when, on the one hand, she proclaims: “With the pure versus applied distinction removed, scientific progress can be defined in terms of the increased capacity to predict, control, manipulate, and intervene in various contexts.” But then, on the other hand, in a footnote she recoils partially: “To be clear, while I think this is a useful rubric for scientific progress, it is not a remotely sufficient account for how one should assess scientific theories.” Other (non-diachronic) theoretical virtues that are complementary to, but less weighty epistemically than, prediction and control also play important roles in theory assessment, she suggests. Consideration of the major non-diachronic theoretical virtues systematized in Sects. 1–5 drives this point home.

How exactly is applicability a diachronic theory trait that is *epistemic* (helping to indicate likely truth) in view of the obvious *pragmatic* orientation of technological application? Agazzi observes that some technological projects “are designed or projected in advance, as the concrete application of knowledge provided by a given science or set of sciences” ([Agazzi 2014, p. 308](#)). If a project of this kind actually works as predicted, then this reinforces our commitment to the theory base that helped guide such action in the world. Agazzi further notes:

The predictions ‘contained’ in the project actually are the predictions made by the scientific theories which have permitted the proposal of the complex *noema* that constitutes the project, and contains not only prescriptions as to the way of realising the structure of the machine but also as to its functioning. This functioning is something that happens; it is a state of affairs that constitutes a confirmation of the theories used in projecting the machine (p. 309).

Although Agazzi’s scientific realism overstates the epistemic reach of applicability, his analysis is illuminating:

A mature science is a science that has given rise to a significant technology. This means, for example, that we can provisionally admit certain theories that are

'empirically adequate,' without admitting their truth as van Fraassen says, until we have significant predictions confirming them. This fact (especially in conjunction with other 'virtues' discussed in the literature) already justifies attributing truth and ontological reference to them, but the existence of technological applications is the last decisive step that assures that they have been able to adequately treat those aspects of reality they intended to treat. (p. 310).

To nuance Agazzi's insightful but inflated epistemic role for applicability, we can observe that this theoretical virtue is not commonly operative in certain scientific domains. For example, scientific theories of "how things originated" (history of nature) lead to fewer technological applications than scientific theories of "how things work." Part of the reason for the infrequent applicability of origins theories is the smaller role that experimentally controlled prediction plays in such theorization. For example, much of the data that allows us to reconstruct the *history* of earth's surface is collected by means of passive field observations, rather than by laboratory experiments that make precise predictions and technological control more feasible.

Hacking, in his own loud italics, expressed an entity realism (not theory realism) that need not be adopted to recognize how his classic work on "representing and intervening" (especially when read in the light of Agazzi) is supportive of applicability as a theoretical virtue of considerable epistemic standing.

We are completely convinced of the reality of electrons when we regularly set out to build—and often enough succeed in building—new kinds of device that use various well understood causal properties of electrons to interfere in other more hypothetical parts of nature (1983, p. 265).

Applicability deserves recognition as one of the major theoretical virtues, which is why I have introduced it and mapped its location in my systematization. Discerning the existence and nature of neglected theoretical virtues and increasing our understanding of the traits of a good theory reveal the importance of systematically studying and classifying the theoretical virtues.

7 The big picture that emerges from this theoretical virtue systematization

There are at least twelve major theoretical virtues and they are best classified into four categories: evidential, coherential, aesthetic, and diachronic.¹² The evidential virtues, which are about how well theoretical components *correspond* to events and regularities in the world, are to be distinguished from the coherential virtues, which pertain to how well theoretical components *fit together*. The aesthetic theoretical virtues possess an aesthetic shape (*fittingness*) that is quite different from the logical-conceptual fit of the coherential virtues. This deep divide is also reflected by the greater difficulty of establishing the epistemic value of aesthetic virtues compared to coherential virtues.

¹² Carrier (2013, p. 2550) documents several other unsystematic lists of theoretical virtues, none of which mentions a virtue beyond the twelve major virtues in my systematization. Of course alternative names for individual virtues populate the literature, as Carrier and I have noted.

Initially one might surmise, on account of their more a priori status, that the coherent virtues and aesthetic virtues are more closely related to each other than either is related to the evidential virtues. However, all four classes of theory virtues involve the human intellect in at least some degree of both a priori and a posteriori reasoning. Thus such distinctions are not among the primary criteria for theoretical virtue taxonomic ranking. Let us develop this point further. Recognition of the evidential virtue of causal adequacy in a particular theory is not just an a posteriori mental operation. There are a priori components too, as the philosophical literature on causation indicates. The same goes for the evidential virtue of explanatory depth. Multiple depth measures have been explored, and they involve various philosophical assumptions. The depth measure I emphasized most, because it is a leading one, is the number of counterfactual questions that a theory's law-like generalizations answer regarding the item being explained. Even evidential accuracy, the most a posteriori evidential virtue, involves theorists in a priori reasoning about what sorts of things potentially *count* as evidence (McCain 2014).

Furthermore, despite the disparity in epistemic value (as typically judged) between the evidential virtues (high value) and the aesthetic virtues (zero or modest value), the aesthetic virtues of simplicity and unification are complementary artistic styles of how theoretical content relates to evidence, and thus are significantly entangled with the evidential virtues. This also helps support the conclusion that the aesthetic virtues are not merely pragmatic (e.g., simple theories are easier to work with). Indeed, the aesthetic theoretical virtues (at least simplicity and unification) might also have modest epistemic credentials (at least extrinsically, and maybe even intrinsically). Additionally, as Douglas (2013) would agree, simplicity and unification might have epistemic value grounded in their correlation (at least in some mathematical model selection methods) with higher predictive success, which is central to the diachronic virtues. However, despite their curious secondary relationships to the evidential and diachronic virtues, simplicity and unification are more fundamentally aesthetic theory virtues that are (perhaps) epistemically significant in multiple ways. Note that my systematization does not deal much with the pragmatic aspects of theory traits (e.g., simple theories are easier to work with, as addressed in Douglas' work). This is because I have aimed primarily at understanding and classifying the theoretical virtues as traits that are *broadly epistemic in theory evaluation*, rather than investigate the pragmatic aspects of theory traits that, for example, might help maximize research efficiency, which certainly is valuable.

So the first three classes of theoretical virtues are arranged in order of decreasing epistemic weight (with the possibility that all or some of the aesthetic virtues carry zero epistemic weight) and each class contains at least three closely related virtues. The fourth class of theoretical virtues—the diachronic—entails a temporal dimension that is missing in the previous classes. The diachronic virtues can only be instantiated *after* a theory's initial formulation—when it has had opportunity to be tested, elaborated, and applied. Durability, fruitfulness, and applicability build upon the initial theory assessment process governed by the non-diachronic virtues (the evidential, coherent, and aesthetic theoretical virtues). The cumulative result, when successful, is a mature theory with even greater epistemic value than an infant theory that has not yet had the opportunity to show whether it will possess the diachronic theoretical virtues.

Each virtue class contains at least three virtues that sequentially follow a repeating pattern of progressive disclosure and expansion. Within the evidential theoretical virtues we observed a progression from achieving a basic evidential fit, to identifying an adequate causal story, and finally to deepening the explanatory account of the evidence in multiple possible ways. The coherential theoretical virtues express how well theoretical components fit together in an increasingly expansive manner. Internal consistency refers to adherence to basic logical rules. Internal coherence pertains to how theoretical components are coordinated into an intuitively plausible whole. Universal coherence is about how a theory “sits well” within one’s total knowledge. The aesthetic theoretical virtues follow a similar pattern: the basic aesthetic property of beauty comes first, followed by two epistemically-enhanced special cases of aesthetic properties: simplicity and unification—constituting, respectively, an inward looking and an outward looking style of theory-evidence aesthetic relation.

The diachronic virtues build upon the first three virtue classes in another three-stage disclosure-expansion pattern that brings the entire epistemic dance to its climax. Durability is instantiated as a theory passes tests in a series of encounters with the world, especially by successful prediction and plausible accommodation of new evidence. Fruitfulness discloses a theory’s resourcefulness yet further through innovation—stimulating additional discovery by successful novel prediction, unification, non ad hoc theoretical elaboration, and other means. At last, applicability expands the epistemic accountability of a theory into the final frontier: the vast domain of practical action in the world. Indeed, the diachronic theoretical virtues provide an ongoing and epistemically intensified means of theory development that complements the non-diachronic virtue assessment process that begins in a theory’s original construction.

This curious repeating disclosure-expansion pattern within each virtue class, which undergoes an intensified repetition in the diachronic class, suggests that theory virtues offer a coordinated and cumulative means of achieving our broadly epistemic aims. This conclusion is also consistent with several other virtue relations analyzed above, such as simplicity and unification as mirror image aesthetic siblings that specify certain theory-evidence relations. For this reason and others, evidential accuracy (empirical fit), according to my systematization, is not a largely isolated trait of good theories, as some (realists and antirealists) have made it out to be (McMullin 2014; Van Fraassen 1980). Rather, it bears multifaceted relations, constituting significant epistemic entanglements, with other theoretical virtues. Indeed, each theoretical virtue is best understood in its mutual relations to the other theoretical virtues. That is why an adequate taxonomy of theoretical virtues is so important.

8 Conclusion

My taxonomic project is both descriptive and prescriptive. I described what past theorists have recognized as the admirable traits of some of the most widely held scientific theories. I also described how philosophers have characterized, but rarely systematized, the theoretical virtues. I offered reasons for accepting my own systematization as superior to previous attempts, which is a prescriptive endeavor. A superior classification system more adequately characterizes certain entities and their mutual relations.

Finally, my systematization offers resources for future prescriptive studies of how theoretical virtues might have a coordinated role in research across the disciplines—with allowance for discipline specific modification.

An informal and flexible logic of theory choice is in the making here. Efforts to formalize each of the major theoretical virtues might benefit from my informal umbrella account. Moreover, logic textbook authors who treat abduction and “inference to the best explanation” would do well to employ a more systematic approach to the criteria for what counts as a “best explanation.” Random lists of a half dozen sample theoretical virtues are inadequate for this branch of logic. Finally, a systematization of the major theoretical virtues may have the potential to guide fruitful collaboration among logicians, epistemologists, and practitioners within specific disciplines such as artificial intelligence (Flach and Kakas 2000).

Ernan McMullin (1924–2011) concluded his final essay on theoretical virtues with this:

The most important discovery in the history of science to date has been the manner in which that activity itself should be carried on and what expectations should guide it. The expectations I have called “theory virtues” have helped to shape it well (2014, p. 570).

Theory virtues, especially as more comprehensively and precisely systematized in the present essay, might (with appropriate contextualization) assist academic endeavor across the disciplines—not just in the natural sciences. However, McMullin’s last word on this subject rightly reminds us that a series of philosophically inclined scientists and philosophers of science have led the way in recognizing, refining, and more skillfully utilizing these rational tools in theory choice.

My trail of discovery is worth highlighting as we conclude. I began with a desire to understand all the major theoretical virtues that are potentially operative in any discipline. My initial hunch that most reflection on theory virtues would be found in the history and philosophy of science (my own field) proved correct. As I waded through theory virtue literature I merely listed and characterized the virtues as I encountered them—typically scholars treat them in arbitrary order. Later I grouped the virtues by mutual affinities, trying several arrangements based on different criteria. I then compared the few previous systematizations (especially Douglas 2013; Mackonis 2013; McMullin 2014) with the patterns revealed by my own work. To guard against shoe-horning the virtues into an idiosyncratically pleasing pattern, I often stepped into the conceptual shoes of other scholars, both the few systematizers and the many others writing on one or more of the individual theoretical virtues. I invite others to retrace my steps, and explore further, so that we may better recognize a good theory when we see one.

Acknowledgements For helpful comments, I thank James Barham, J. Thomas Bridges, Amanda Bryant, Travis Dickinson, Paul Gould, and three anonymous referees.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agazzi, E. (2014). *Scientific objectivity and its contexts*. Cham: Springer.
- Axtell, G. (2014). Bridging a fault line: On underdetermination and the ampliative adequacy of competing theories. In A. Fairweather (Ed.), *Virtue epistemology naturalized* (pp. 227–246). Cham: Springer.
- Beebe, J. R. (2009). The abductivist reply to skepticism. *Philosophy and Phenomenological Research*, 79(3), 605–636.
- Benovsky, J. (2013). Philosophical theories, aesthetic value, and theory choice. *The Journal of Value Inquiry*, 47(3), 191–205.
- Boudry, M. (2013). The hypothesis that saves the day: Ad hoc reasoning in pseudoscience. *Logique et Analyse*, 223, 245–258.
- Breitenbach, A. (2013). Aesthetics in science: A Kantian proposal. *Proceedings of the Aristotelian Society*, 113, 83–100.
- Carrier, M. (2013). Values and objectivity in science: Value-ladenness, pluralism and the epistemic attitude. *Science & Education*, 22(10), 2547–2568.
- Cleland, C. E. (2011). Prediction and explanation in historical natural science. *British Journal for the Philosophy of Science*, 62(3), 551–582.
- Collins, R. (1994). Against the epistemic value of prediction over accommodation. *Noûs*, 28(2), 210–224.
- DeLancey, C. (2011). Does a parsimony principle entail a simple world? *Metaphysica*, 12(2), 87–100.
- Dirac, P. A. M. (1980). The excellence of Einstein's theory of gravitation. In M. Goldsmith, et al. (Eds.), *Einstein: The first hundred years* (pp. 41–46). Oxford: Pergamon Press.
- Douglas, H. (2009). Reintroducing prediction to explanation. *Philosophy of Science*, 76(4), 444–463.
- Douglas, H. (2013). The value of cognitive values. *Philosophy of Science*, 80(5), 796–806.
- Douglas, H. (2014). Pure science and the problem of progress. *Studies in History and Philosophy of Science Part A*, 46, 55–63.
- Douglas, H., & Magnus, P. D. (2013). State of the field: Why novel prediction matters. *Studies in History and Philosophy of Science Part A*, 44(4), 580–589.
- Flach, P. A., & Kakas, A. C. (2000). *Abduction and induction*. Dordrecht: Kluwer.
- Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *The British Journal for the Philosophy of Science*, 45(1), 1–35.
- Frankel, H. R. (2012). *The continental drift controversy* (Vol. 4). Cambridge: Cambridge University Press.
- Friederich, S., et al. (2014). Philosophical perspectives on ad hoc hypotheses and the higgs mechanism. *Synthese*, 191(16), 3897–3917.
- Frisch, M. (2016). Peter Vickers understanding inconsistent science. *The British Journal for the Philosophy of Science*, 67(3), 913–918.
- Glynn, I. (2010). *Elegance in science: The beauty of simplicity*. Oxford: Oxford University Press.
- Greene, M. T. (2015). *Alfred Wegener*. Baltimore: Johns Hopkins University Press.
- Hacking, I. (1983). *Representing and intervening*. Cambridge: Cambridge University Press.
- Harker, D. (2008). On the predilections for predictions. *The British Journal for the Philosophy of Science*, 59(3), 429–453.
- Hitchcock, C., & Woodward, J. (2003). Explanatory generalizations, part II: Plumbing explanatory depth. *Noûs*, 37(2), 181–199.
- Houkes, W. (2009). The nature of technological knowledge. In *Philosophy of technology and engineering sciences* (pp. 309–350). Amsterdam: North-Holland.
- Huemer, M. (2009). When is parsimony a virtue? *Philosophical Quarterly*, 59(235), 216–236.
- Hunt, J. C. (2012). On ad hoc hypotheses. *Philosophy of Science*, 79(1), 1–14.
- Jardine, N., & Kepler, J. (1984). *The birth of history and philosophy of science*. Cambridge: Cambridge University Press.
- Kelly, K. T. (2011). Simplicity, truth, and probability. In P. Bandyopadhyay & M. Forster (Eds.), *Handbook on the philosophy of statistics* (pp. 983–1024). Dordrecht: Elsevier.
- Kelly, K. T., Genin, K., & Lin, H. (2016). Realism, rhetoric, and reliability. *Synthese*, 193(4), 1191–1223.
- Kepler, J. (1981). *Mysterium cosmographicum* (A. M. Duncan, Trans.). New York: Abaris.
- Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, 48(4), 507–531.
- Knuutila, T., & Loettgers, A. (2011). Causal isolation robustness analysis: The combinatorial strategy of circadian clock research. *Biology & Philosophy*, 26(5), 773–791.
- Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In T. S. Kuhn (Ed.), *The essential tension* (pp. 310–339). Chicago: University of Chicago.

- Kuhn, T. S. (1993). Afterwords. In P. Horwich (Ed.), *World changes: Thomas Kuhn and the nature of science* (pp. 311–341). Cambridge: MIT Press.
- Laudan, L. (1984). *Science and values*. Berkeley: University of California Press.
- Laudan, L. (2004). The epistemic, the cognitive, and the social. In P. K. Machamer & G. Wolters (Eds.), *Science, values, and objectivity* (pp. 14–23). Pittsburgh: University of Pittsburgh Press.
- Levinson, J. (2003). Philosophical aesthetics: An overview. In J. Levinson (Ed.), *The Oxford handbook of aesthetics* (pp. 3–24). Oxford: Oxford University Press.
- Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). London: Routledge.
- Longino, H. (1996). Cognitive and non-cognitive values in science: Rethinking the dichotomy. In L. H. Nelson & J. Nelson (Eds.), *Feminism, science, and the philosophy of science* (pp. 39–58). Dordrecht: Kluwer.
- Mackonis, A. (2013). Inference to the best explanation, coherence and other explanatory virtues. *Synthese*, 190(6), 975–995.
- Mayo, D. (2014). Some surprising facts about (the problem of) surprising facts. *Studies in History and Philosophy of Science Part A*, 45, 79–86.
- McAllister, J. W. (1996). *Beauty and revolution in science*. Ithaca: Cornell University Press.
- McCain, K. (2014). *Evidentialism and epistemic justification*. New York: Routledge.
- McCartney, M. (2015). Introduction. In S. Lawrence & M. McCartney (Eds.), *Mathematicians and their gods* (pp. 1–20). Oxford: Oxford University Press.
- McMullin, E. (1983). Values in science. In P. Asquith & T. Nickle (Eds.), *PSA 1982: Proceedings of the 1982 biennial meeting of the philosophy of science association* (Vol. 2, pp. 3–28). East Lansing: Philosophy of Science Association.
- McMullin, E. (1987). Explanatory success and the truth of theory. In N. Rescher (Ed.), *Scientific inquiry in philosophical perspective* (pp. 51–73). Lanham: University Press of America.
- McMullin, E. (1993). Rationality and paradigm change in science. In P. Horwich (Ed.), *World changes: Thomas Kuhn and the nature of science* (pp. 55–78). Cambridge: MIT Press.
- McMullin, E. (1996). Epistemic virtue and theory-appraisal. In I. Douven & L. Horsten (Eds.), *Realism in the sciences* (pp. 1–34). Leuven: Leuven University Press.
- McMullin, E. (2012). Values in science 1983. *Zygon*, 47(4), (pp. 686–709). (Reprint of McMullin, 1983).
- McMullin, E. (2014). The virtues of a good theory. In M. Curd & S. Psillos (Eds.), *The Routledge companion to philosophy of science* (pp. 561–571). New York: Routledge.
- Montano, U. (2013). Beauty in science: A new model of the role of aesthetic evaluations in science. *European Journal for Philosophy of Science*, 3(2), 133–156.
- Montano, U. (2014). *Explaining beauty in mathematics*. Dordrecht: Springer.
- Nerlich, G. (1979). What can geometry explain? *The British Journal for the Philosophy of Science*, 30(1), 69–83.
- Nolan, D. (1999). Is fertility virtuous in its own right? *The British Journal for the Philosophy of Science*, 50(2), 265–282.
- Radder, H. (2009). Science, technology and the science-technology relationship. In A. Meijers (Ed.), *Philosophy of technology and engineering sciences, Handbook of the philosophy of science* (Vol. 9, pp. 65–91). Amsterdam: North Holland.
- Reiss, J., & Sprenger, J. (2014). Scientific objectivity. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/scientific-objectivity>. Accessed 22 July 2016.
- Schindler, S. (2014). Novelty, coherence, and Mendeleev's periodic table. *Studies in History and Philosophy of Science Part A*, 45, 62–69.
- Segall, R. (2008). Fertility and scientific realism. *The British Journal for the Philosophy of Science*, 59(2), 237–246.
- Smith, G. E. (2010). Revisiting accepted science: The indispensability of the history of science. *Monist*, 93(4), 545–579.
- Smith, G. E. (2014). Closing the loop: Testing Newtonian gravity, then and now. In Z. Biener & E. Schliesser (Eds.), *Newton and empiricism* (pp. 262–351). New York: Oxford University Press.
- Soler, L., et al. (Eds.). (2012). *Characterizing the robustness of science*. Dordrecht: Springer.
- Sober, E. (2015). *Ockham's razors*. Cambridge: Cambridge University Press.
- Steel, D. (2010). Epistemic values and the argument from inductive risk. *Philosophy of Science*, 77(1), 14–34.
- Strevens, M. (2008). *Depth: An account of scientific explanation*. Cambridge: Harvard University Press.
- Swinburne, R. (1997). *Simplicity as evidence of truth*. Milwaukee: Marquette University Press.

- Thagard, P. R. (1978). The best explanation: Criteria for theory choice. *The Journal of Philosophy*, 75(2), 76–92.
- Van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.
- Vickers, P. (2013). *Understanding inconsistent science*. Oxford: Oxford University Press.
- Vincenti, W. G. (1990). *What engineers know and how they know it*. Baltimore: Johns Hopkins University Press.
- Weisberg, M. (2006). Robustness analysis. *Philosophy of Science*, 73(5), 730–742.
- Williams, M. B. (1973). Falsifiable predictions of evolutionary theory. *Philosophy of Science*, 40(4), 518–537.
- Winther, R. G. (2009). Prediction in selectionist evolutionary theory. *Philosophy of Science*, 76(5), 889–901.
- Ylikoski, P., & Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical Studies*, 148(2), 201–219.
- Zangwill, N. (2001). *The metaphysics of beauty*. Ithaca: Cornell University Press.
- Zangwill, N. (2014). Aesthetic judgment. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/aesthetic-judgment>. Accessed 23 July 2016.