5-2017

The Beauty of Understanding: Aesthetic Methods of Theory Evaluation

Devon Craig Bryson
University of Tennessee, Knoxville, dbryson6@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Devon Craig Bryson entitled "The Beauty of Understanding: Aesthetic Methods of Theory Evaluation." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Philosophy.

David Reidy, Major Professor

We have read this dissertation and recommend its acceptance:

Jon Garthoff, David Palmer, Allen Dunn

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

The Beauty of Understanding:
Aesthetic Methods of Theory Evaluation

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Devon Craig Bryson
May 2017

ABSTRACT

Philosophers use a variety of methods to evaluate theories, theories that are sources of greater understanding. My dissertation argues that judgments of beauty are a justified part of how we evaluate theories. That is, I argue that beauty is part of what makes a philosophical theory better. I reach this conclusion by analyzing two powerful and popular methods of theory evaluation: reflective equilibrium and simplicity. The literatures on both reflective equilibrium and simplicity clarify how these methods work and why they are justified methods of theory evaluation. But I argue that the going accounts of reflective equilibrium and simplicity have gaps remaining. Both methods rely on judgments that are unexplained. Reflective equilibrium requires judgments of coherence and simplicity requires judgments of simplicity. Yet the going accounts give no explanation of how to make these judgments. I argue that these gaps are best filled by identifying judgments of coherence and judgments of simplicity as species of judgments of beauty. Judgments of coherence and simplicity should be identified as species of judgments of beauty because they share a special character as unprincipled, yet genuine, judgments. That is, all three kinds of judgment are not made by reference to principles, and yet reasonable, non-arbitrary judgments are possible. This identification completes the accounts of reflective equilibrium and simplicity because it explains how we make judgments of coherence and simplicity, despite lacking a principled account of those judgments. This means that two powerful and popular methods of theory evaluation do in fact use judgments of beauty to identify better theories. I conclude by arguing that using judgments of beauty to identify better theories is justified because of the fundamental role these judgments play in guiding theory evaluation.

TABLE OF CONTENTS

CHAPTER ONE

THEORIES

Section 1: A Theory of Theories

When we contrast two different answers to a single philosophical question we usually contrast two different theories. This is clearest at the highest levels of generality. Suppose someone asks us, "What makes an action moral?" In our answer we will likely contrast, say, the theory of utilitarianism with the theory of Kantian ethics. But even at lower levels of generality, our answers will still be along these same lines. Suppose the person asks instead, "What makes promise keeping moral?" or if she asks, "In situation Q, with circumstances X, Y, and Z, should I keep my promise?" Our answer is still likely to be to contrast some theories and then proceed to apply them to the specifics. The point is a simple one: Theories are what we are usually looking for in seeking answers to philosophical questions. Whether our question is about morality, knowledge, action, justice, free will, or language, the various answers take the form of theories. These answers all compete to be the best theory of the given phenomenon. The various –isms that populate philosophy are such theories. When we teach students, we contrast theories; when we defend our views, we defend a particular theory; when we work on a question, we build up competing theories to see which one comes out better.

There is nothing mysterious about this point. Philosophical questions tend to be questions that seek explanations for the phenomena that matter most to us. Take again the question, "What makes an action moral?" It is simply the wrong kind of answer to respond, "Promise keeping makes an action moral." The question was not seeking a mere list of moral actions (or types of moral actions); it was seeking an explanation of the nature of moral actions. A theory is the proper kind of answer to most philosophical questions because it purports to provide the sought explanation.[1] A mere list of facts relevant to the phenomenon can never provide an explanation in the way that a theory of the phenomenon can. We can take this one step further: An explanation of the phenomenon is sought because we seek to understand the phenomenon. So, if we ultimately want to understand the things that matter most to us, we want explanations of those things that are satisfying to our understanding. And theories are the right kind of thing for giving us such explanations. Theories, explanation, and

---

[1] There is, of course, some philosophical debate about what precisely *explanation* is. This debate is a debate over what theory of explanation is the best, what best explains the phenomenon of explanation. See Nozick (1981).

understanding are thus critical notions for articulating the purpose and method of philosophy (and science, for that matter): The best theory of a phenomenon is sought so that we can have the best explanation of that phenomenon so that we can understand the phenomenon as best we can.

So, if a theory is the proper form for an answer to a philosophical question, then philosophers will want to know what a theory is. That is, we want a theory of theories. This dissertation develops a (partial) theory of theories. The first part of a theory of theories is to say what theories are. What theories are precisely is, of course, subject to debate, particularly in the philosophy of science.[2] It is beyond the scope of this project to venture into that debate. For our purposes, the answer need not be complicated. In the theory of theories that I develop, I understand a "theory" to be a "system of ideas, especially one built around general principles, that purports to explain something." I have already gestured at the connection between theories and explanation, so for now let's focus on the first part: a theory is a "system of ideas, especially one built around general principles". We could of course ask what a "system" is or what "general principles" are. The theory of theories that I develop has more to say about each of these. But for now, when we're just trying to get a fix on this definition, it helps to turn to some examples.

Philosophy is populated by many clear examples of theories. For any given philosophical subject matter, there are usually a few different theories (or families of theories) of that subject matter. For the subject of morally right action, there are (to name but two) the theories of utilitarianism and Kantian ethics. The theory of utilitarianism is built around principles of reducing harm and producing more good. Ideas of pleasure and pain, maximization, and impartiality often play central roles. These principles, as well as the attendant systems of ideas that fill them out, purport to explain what makes an action morally right. In contrast, the theory of Kantian ethics is built around principles of giving respect to all people. Ideas of personality, dignity, and universality often play central roles. Again, these principles, as well as the attendant systems of ideas that fill them out, purport to explain what makes an action morally right. These descriptions are, of course, purposefully vague because there are many variations that refine these theories in various ways. (They are thus more properly described as families of theories.) These theories can be properly said to be competing theories because they use substantially different systems of ideas to explain the same target phenomenon.[3]

---

[2] See Nozick (1981) and Ramsey (1931).

[3] There are, of course, difficulties in individuating theories and in marking out what constitutes a *different* theory. There are also difficulties in individuating what counts as the *same* target phenomenon and therefore in demarcating which theories are *actually* in competition. Often two theories can *seem* to be competing, but showing that they actually explain *different* phenomenon dispels this illusion. But these difficulties can be set-aside for the present purpose, the rough individuations are clear enough.

The same situation—there being multiple families of competing theories—holds true for virtually every live philosophical question. An incomplete list would include: In free will debates: compatibilism vs. incompatibilism. In epistemology: competing solutions to the Gettier problem. In philosophy of action: noncausal theories of reasons-explanation vs. causal theories of reasons-explanation. In political philosophy: libertarianism vs. Rawls' theory. All of these are systems of ideas, usually built around general principles, that purport to explain a target phenomenon in a more satisfying way than their rivals. Thus, these are all theories by the definition given above.

There are familiar examples of theories in the natural sciences as well. To use the well-worn example: Ptolemy's theory vs. Copernicus' theory vs. Kepler's theory. These are competing theories because they purport to explain the same target phenomenon—the motions of the heavenly bodies—using different systems of principles and ideas. The core principles of the Ptolemaic system are geocentrism and the circular movement of the celestial bodies. Copernicus' theory kept the principle of the circular movement, but adopted the principle of heliocentrism. Kepler's theory preserves the principle of heliocentrism, but adopts the elliptical movement of the celestial bodies. Of course, each theory is filled out from these principles in different ways to account for the same data: the observations of the movements of the heavens. Thus, we see the same pattern in science as in philosophy: the goal is to explain a target phenomenon, so mere lists of facts (or data in the case of science) are not satisfying. Instead, a theory—a system of principles and ideas—is needed to explain the phenomenon. Different theories compete in science to be the better theories of the target phenomenon. Darwinian vs. Lamarckian evolutionary theories provide another example: they purport to explain the same phenomenon—the development of the diversity of species—using a different system of principles and ideas. Classical mechanics vs. quantum mechanics provide yet another example: the same phenomenon—the motion of bodies—is explained using a different system of principles and ideas.

But we don't even need to look to the academic realm to see easy examples of theories. Indeed, the academic examples I have given thus far are all relatively grand theories, and so are somewhat unrepresentative examples. Theories need not aspire to such breadth of explanation. Explanations of mundane phenomena are also good cases of theories and are more common. Here is an example of an extremely rudimentary theory, which I will refer to throughout this dissertation: "Mr. Boddy was found, dead, in the library. Mr. Boddy's blood was found on the candlestick. Colonel Mustard was found in the library holding the candlestick. Colonel Mustard stood to inherit Mr. Boddy's estate. We take the basic rules of probability and human behavior as our principles. Thus, Colonel Mustard killed Mr. Boddy." This theory is a system of ideas that purports to explain a phenomenon: the death of Mr. Boddy. It explains this phenomenon by asserting that Colonel Mustard killed Mr. Boddy, which is in turn supported by Colonel Mustard's motive and means to kill Mr. Boddy (the

principles of probability and human behavior, though in the background, form the backbone of this theory). This same target phenomenon (the death of Mr. Boddy) could also perhaps be explained by a different theory, one that involved Colonel Mustard being framed. These different theories would be competing theories and we would have to work to see which was the better theory. In this way, we use systems of ideas—theories—to explain mundane phenomena of all sorts.

This gamut of examples shows that, by the definition above, theories are not mysterious or arcane objects. Theories come in various levels of breadth and depth: in science and philosophy, we sometimes focus on grand theories (theories that attempt to explain much with very little), but narrow theories are no less theories. Any time there is an ambition to explain—an extremely common ambition for humans—a system of ideas is deployed to meet the task. Such a system is a theory. These examples also point to what the proper opposite of a theory is: an unsystematic set of ideas (perhaps related to the target phenomenon) that does not explain the target phenomenon. An unsystematic set of ideas is something like a mere list of ideas. In philosophy, this is something like a mere list of claims related to the phenomenon: X, Y, and Z, are instances of moral action (or knowledge or justice, etc.). In science, this is something like a mere list of observations. In explanations of mundane phenomena, this is something like a mere list of facts about the phenomenon in question. It is, of course, sometimes difficult to tell when a list becomes organized or systematized enough to become more than a mere list, but the rough distinction is clear enough.

But a theory of theories should not only explain what theories are. In building a theory of theories we also want to know what the purpose of theories is. As I have already emphasized, and as is built into the definition I gave above, the core purpose of theories is to explain phenomena. To explain something is, roughly, to make that thing clearer by revealing or illuminating relevant ideas.[4] That is to say, to explain to something is, roughly, to decrease our sense of mystery about that thing. Explaining X requires pointing to some other thing, which illuminates X, decreases our sense of mystery about X. Often explanation can be framed as an answer to a "why" question: "Why does morality have the structure it does? Why do the heavenly bodies move the way they do? Why is Mr. Boddy dead?" To answer these questions, by reference to something other than the target phenomenon, is to explain the target phenomenon. What is this other thing that we refer to in order to explain a phenomenon? It is a theory. If explanation is a matter of revealing relevant ideas, then theories—which are systems of relevant ideas—are exactly the other thing we must refer to in order to explain something. Theories and explanation are thus tightly linked together concepts. But to continue filling out the purpose of theories, we might press on

---

[4] Again, there are debates in philosophy, particularly in the philosophy of science, over the *precise* nature of explanation. For our purposes here, the rough understanding is sufficient.

and ask: "All well and good that theories provide explanations, but what is the purpose of explanation?"

Many more specific benefits fall out of the definition of explanation given above. Explanation can be used in descriptive domains: this would be an explanation of how a phenomenon actually is or works, such as how the natural world works. Here such an explanation also often provides predictions about the phenomenon as well as guidance in how to navigate the phenomenon. Such predictions and guidance are of practical value to us, so it is no wonder that we should seek out explanations of phenomenon that empower us to make predictions about and that guide our navigation of the phenomenon. Explanation can also be used in the normative domain: this would be an explanation of how a phenomenon should be, such as how a government should be organized. Here such an explanation also often provides a decision procedure or a deliberative framework for working through the relevant normative domain. These more practical upshots of explanation might be called the guiding benefits of explanation. In making a phenomenon clearer by revealing or illuminating relevant ideas, explanation often delivers these guiding benefits for the target phenomenon.

In addition to these guiding benefits, explanation also provides a more purely epistemic benefit: understanding. To understand something is to have a mental grasp of it. And it's intuitive to see how explaining something helps one to grasp it. But the notion of understanding is enjoying something of a renaissance in contemporary philosophy, so we can look to accounts of understanding that move beyond this rough and metaphorical idea of understanding. Perhaps the most complete treatment comes from the work of Jon Kvanvig.[5] He gives an account of understanding as follows: "Understanding requires the grasping of explanatory and other coherence-making relationships in a large and comprehensive body of information. One can know many unrelated pieces of information, but understanding is achieved only when informational items are pieced together by the subject in question."[6] And again: "What is distinctive about understanding has to do with the way in which an individual combines pieces of information into a unified body."[7] Kvanvig is concerned not only with giving an account of the nature of understanding, but also with giving an account of the value of understanding.[8] On that score he concludes: "Understanding is valuable because it is constituted by subjectively justified true belief across an appropriately individuated body of information that is systematized and organized in the process of achieving understanding, and subjectively justified true belief that is systematized in this way is valuable."[9]

---

[5] For Kvanvig's most complete treatment of understanding, see Kvanvig (2003).
[6] Kvanvig (2003), p. 192.
[7] Kvanvig (2003), p. 197.
[8] More specifically, he is concerned to give an account on which understanding is *more* valuable than any proper subset of its constituents. More on this in a moment.
[9] Kvanvig (2003), p. 202.

There is much in Kvanvig's account that is fruitful, but it is beyond the scope of this project to fully explore it. For our purposes, it is enough to glean two things from this account: 1) Understanding is something like "systematic knowledge". "Systematic knowledge" is the clearest, simplest way to put it, but it's not quite accurate. On Kvanvig's account, understanding does not, strictly speaking, entail knowledge. This is because knowledge is vulnerable to Gettier-style undermining, whereas understanding is not.[10] It is thus more accurate to say (as Kvanvig does in the quote above) that understanding is "systematic subjectively justified true belief".[11] Thus, understanding emerges as an epistemic phenomenon of its own kind, distinct from other more familiar epistemic notions: truth, justification, and knowledge. Truth and (subjective) justification are constituents of understanding (at least on Kvanvig's account) and knowledge is surely a close cousin of understanding. But understanding is distinct from all of these, and knowledge is neither necessary nor sufficient for understanding (again, on Kvanvig's account). 2) Understanding is valuable, among our most valuable epistemic achievements. This point is far more significant for Kvanvig's purposes. His project is to address the Meno problem: the question of whether knowledge is more valuable than any proper subset of its constituents.[12] After many examinations of potential sources of value for knowledge, Kvanvig concludes that knowledge cannot survive the Meno problem. That is, he concludes that knowledge is not more valuable than any proper subset of its constituents. This is of course compatible with knowledge being valuable (indeed, of great value), as Kvanvig is at pains to stress. It's just that—contrary to what one might have hoped—the value of knowledge is exhausted by some proper subset of its parts (namely, on Kvanvig's account, the elements of knowledge minus the element that is necessary to insulate knowledge from Gettier phenomena). In contrast to all this, Kvanvig develops his account of understanding to show that understanding can survive the analogue of the Meno problem. That is, understanding, unlike knowledge, is more valuable than any proper subset of its constituents. This gives it (on Kvanvig's account) a unique kind of epistemic value.

---

[10] A Gettier case is one in which one has justified true belief, but not knowledge. For example, suppose you read a stopped clock (without knowing that it has stopped) at precisely the time when the clock reads X and it *actually is* X. If you then think "The time is now X", then your knowledge of the time will have been Gettier-style undermined. You are justified in believing what clocks tell you about the time, and your belief is true, but you do not *know* what time it is. See Gettier (1963).

[11] For ease of expression, I continue to gloss understanding as "systematic knowledge" but this point should be kept in mind.

[12] This question—Is knowledge more valuable than any proper subset of its parts?—is called the *Meno* problem because Plato addresses it in his dialogue *Meno*. It's a problem because it seems like knowledge should be more valuable than any proper subset of its parts, but it is difficult to show this. See Plato, *Meno*. See also Pritchard and Turri (2014).

Given this account of understanding, it is clear how explanation can contribute to understanding. If explanation is making a phenomenon clearer by revealing relevant ideas, then such illumination of relevant ideas is critical to having systematic knowledge of the phenomenon. And, what's more, it's clear that we have come full circle: The proper object of understanding is a "systematized body of knowledge" or a "comprehensive body of information." This echoes the definition of "theory" that I gave above. The chain from theory to explanation to understanding is thus clear. A theory is a system of ideas that purports to explain some target phenomenon. Such a system is an abstract object, like an equation or a schematic or a piece of music. We are able to use such abstract objects to explain their target phenomena, to make those phenomena clearer by revealing relevant ideas. Those relevant ideas are precisely those systematized in the theory. In so explaining their target phenomena, theories engender in us understanding. This is because in having the target phenomenon explained to us, we come to have systematic knowledge of it. Understanding is precisely such an epistemic achievement, having systematic knowledge of a subject. Such systematic knowledge is, loosely speaking, a kind of doxastic object, the doxastic equivalent of the abstract theory. By the process of explanation, we are able to transfer the systematicity of the theory to the systematicity of our knowledge. Systematicity is, thus, a critical notion in this chain and I return to this notion in the final chapter. Finally, whether or not one accepts the details of Kvanvig's account of the value of knowledge (I certainly don't mean to fully endorse them), it is clear enough that understanding is of great epistemic value. This means that this chain of concepts I have developed—theory, explanation, understanding—can satisfactorily terminate with understanding. Understanding is of great epistemic value, so to connect theories to understanding is to secure the importance of theories.

Recall that this sketch of understanding and other epistemic benefits of explanation was part of our efforts to discover the purpose of explanation and thus the purpose of theories. This was in the service of developing a theory of theories. But a theory of theories does not only need to address what theories are and the purpose of theories. These are not the only things about theories that matter to us. We also want to know what makes a theory better or worse or an explanation of how we evaluate theories. Analogously, we want to know what makes an explanation better or worse and what makes for better understanding.[13] Developing an account of what makes a theory better or worse should, in turn, guide us with respect to what makes explanations and understanding better. On the other hand, we might proceed in the opposite direction: An account of what makes a person's understanding of a phenomenon

---

[13] "Inference to the best explanation" is a common method for identifying better theories. But when used flat footedly it doesn't tell us much: Of course we think we should believe the best explanation, but the whole question is, "Which explanation *is* the best?" To answer that question, we need to know what makes an explanation (or a theory) better or worse.

better or worse should, in turn, guide us with respect to what makes explanations and theories better.[14] But such matters would take us too far afield for the purposes of this dissertation. For the purposes of developing a (partial) theory of theories, I restrict myself to considering what makes a theory better or worse and leave the related concepts for another time. If we can get an account of what makes a theory better or worse, then that in conjunction with the rough accounts of what theories are and the purpose of theories given above yields the beginnings of a complete theory of theories.

Section 2: Methods of Theory Evaluation as Guides to What Makes a Theory Good

It is fairly easy to give a rough account of what theories are and the purpose of theories, but no such rough and ready answer is available when it comes to the question of what makes a theory better or worse, of how we evaluate theories. This is troubling, since this is perhaps the most important part of what a theory of theories could tell us: Since theories are so important to us, we want to know how to identify the good theories from the bad. Recall also that theories are often competing to explain the same phenomenon. Resolution of such competitions between theories (whether in philosophy, science, or the mundane) requires an account what makes a theory better or worse. If we are ever to identify which of two (or more) competing theories are better or more satisfying to our understanding, then we need an account of what makes a theory better or worse. We will need an account, an explanation, of what makes a theory better. This will, in turn, allow us to sort and rank theories as better and worse. Thus, a theory of theories, particularly this part of a theory of theories, is critically important to evaluating theories and thus to explaining phenomena in a way most satisfying to our understanding. So, what makes a theory better or worse, how do we evaluate competing theories?

My strategy for making progress on this question is to consider the methods and criteria we use to evaluate theories. As I said above, it is not as though we have an antecedent idea about what makes theories better or worse. Thus, to satisfactorily answer the question of what makes a theory better or worse, we need to look to something other than our judgments of which theories are good. That is, if we are looking for a theory of theories, we need to look to something other than theories themselves to figure out what makes a theory better or worse. My suggestion is that our practice of theorizing—our methods

---

[14] And if the connections between theories, explanation, and understanding are as tight as I have suggested, then this is probably six of one, half dozen of the other: an account of what makes a theory better or worse automatically brings with it an account of what makes understanding better or worse, and vice versa.

and criteria of evaluating theories—can point us to the theoretical material with which we can answer the question of what makes a theory better or worse. After all, it is not as though our methods and criteria merely recommend themselves as tools to get to what we already know to be good theories. This is because, again, we don't have an antecedent idea of which theories are good.   Rather, we—philosophers and scientists—have developed the practice of using certain methods and criteria to evaluate theories and we are willing to follow the methods where they lead, so to speak, even if they lead us in surprising directions. The result of this practice (we hope) is our achieving better and better theories over time. Thus, understanding this practice—our using methods and criteria to evaluate theories—will yield understanding of what makes a theory better or worse. Understanding the structure of our methods and criteria will reveal to us the qualities of a good theory that our practice is predicated upon. So, what methods and criteria do we use to evaluate theories?

This question is complicated by the fact that there are several stages to evaluating theories as better or worse. I have been speaking thus far as though as though the comparison of competing theories happens in one moment, but this is an oversimplification. We can instead think of the evaluation of theories, distinguishing better and worse theories, as happening in several stages. First, there is the process of constructing theories. When we build new theories to explain a new phenomenon (or to explain an old phenomenon if we are unsatisfied with the going theories) there are several methods and criteria we use to build better theories. When we are building new theories we are most beholden to a criterion like "fit with the data". We want to build a theory that captures and explains the facts we know about the phenomenon to be explained. For a scientific theory, this means comportment with our observations; for a philosophical theory, this means comportment with our best judgments. Thus, we can see that "fit with the data", in its various forms, is part of the explanation of what makes a theory better. If we want to understand what it is for a theory to be better, we will want to understand more about what it is for a theory to "fit with the data".

Second, there is the process of selecting between two (or more) already constructed theories that compete to explain the same phenomenon. In this case, we have two (or more) theories that are already up and going that purport to explain the target phenomenon. Presumably this means that they both do fairly well with respect to "fit with the data", the criterion that was used when the theories were being constructed. How, then, do we choose between them? When we are selecting between theories we are most beholden to more formal criteria of theory evaluation: criteria like simplicity. If there are two theories that have been constructed to capture the data, but one is dramatically simpler, we select that one as the better theory.

This makes it seem like there are hierarchies of methods or criteria of theory evaluation. First, one constructs a theory using "fit with the data"; a theory must satisfy this criterion to get up and going. Then, if there are multiple theories

that pass that bar, we go on to select between them using simplicity as a kind of tiebreaker. But this is oversimplified: We undoubtedly take considerations of simplicity into account even as we are building a theory. We may want to capture the data, but we will strive to do so in as simple a way as possible even when we are building our theories. And, on the other hand, we are also able to select between competing theories by using "fit with the data". Two competing theories may both, in a broad sense, satisfy "fit with the data", but we are still able to evaluate competing theories in terms of how well they satisfy "fit with the data". This is because theories can fit better with the data and we can use this evaluation even at the selection stage. Nevertheless the contrast is helpful: "fit with the data" and simplicity are both important criteria of theory evaluation, but they do seem to occupy different positions in theory evaluation. Understanding these criteria—how they work and how they are related—is one of the goals of this dissertation.

Finally, there is a third stage of theory evaluation. This stage is more mysterious and more difficult to understand. This is the stage of reorientation and is famously described by Thomas Kuhn.[15] In this stage, a theory (or set of theories) that has long been established is overthrown in favor of a dramatically new theory (or set of theories). It is difficult to understand how these dramatic shifts—paradigm shifts—happen. It involves something like seeing the target phenomenon in a new way, interpreting the relevant data in a fundamentally different way such that new kinds of fit with that data are made available. The old theories seem to wear out their explanatory power and the new theories offer new frontiers of understanding the target phenomenon. Although this stage is not the direct object of interest for this dissertation, the conclusions I reach may help shed some light on it. We will return to this at the very end of this dissertation.

This schematic is only rough and the distinctions are not hard and fast. As I already indicated above: we use both "fit with the data" and simplicity as criteria in both the construction and selection phase. And even in the more mysterious phase of reorientation, the criteria of "fit with the data" and simplicity are not abandoned. Nor is this schematic meant to be anywhere near exhaustive. I have only gestured at two examples of the kinds of criteria we use to evaluate theories. Remember, this is in service to a theory of theories, specifically, understanding what makes a theory better or worse. My strategy for making progress on this issue is to analyze some methods and criterion of theory evaluation, which are the tools we use to get to better theories. That is, understanding our practice of evaluating theories will help us understand what makes a theory good. I have laid out this schematic of the different stages of theory evaluation to highlight two examples of such methods and criteria of theory evaluation. In the interest of precision, I will restrict my focus in this dissertation to the two examples I have raised thus far: simplicity and "fit with the data" (which, as we will see, is closely related to Rawls' reflective equilibrium,

---

[15] See Kuhn (1962).

which is one of the objects of my analysis). I have distinguished "fit with the data" and simplicity as occupying different roles—theory construction and theory selection, respectively. I will not, however, continually highlight this difference. For ease of expression, I will refer to them both as being criteria of theory evaluation simpliciter. But the fact that "fit with the data" and simplicity occupy different roles in theory evaluation should be kept in mind throughout this dissertation.

Our goal is to understand what makes a theory better. The strategy for understanding this is to analyze methods and criteria of theory evaluation. This is because we use methods and criteria of theory evaluation to get us to better theories. But using them will only help us get to better theories if they are justified methods and criteria of theory evaluation.[16] A justified method for choosing better theories would provide some procedure to go through the outcome of which would be a better theory. A justified criterion for choosing theories would specify some property of theories and a theory that possessed that property would be a better theory.[17] The history of philosophy and science is full of candidate methods and criteria for choosing theories: methods and criteria for choosing theories that purport to be justified, to actually identify better theories. "The scientific method" is a candidate method of evaluating scientific theories, and "fit with the data" is a candidate criterion for evaluating scientific theories (and these both have an analogue in evaluating philosophical theories).[18] Such candidate methods and criteria for choosing better theories can also be framed as "theoretical virtues": the virtues that a theory can have. If being the product of a particular method or having a particular property is a genuine theoretical virtue, then we have identified a justified method or criterion of choosing better theories. There are many candidate theoretical virtues, that is, many candidate methods and criteria for choosing better theories. A theory of theories, in explaining what makes a theory better or worse, would identify candidate methods and criteria of theory evaluation and give reasons why those methods and criteria are justified methods and criteria. That is, it would give an explanation of why those methods and criteria deliver better theories. An explanation of why those methods and criteria deliver better theories will continue to fill out our theory of theories. Remember, my strategy for answering the question "What makes a theory better?" is to try to understand our practices for evaluating theories. But in order for that strategy of answering our question to be

---

[16] For more on justification and theories, see Alston (2005), Kvanvig (2003), Nozick (1981), Sosa (1980), and Zagzebski (2009).

[17] Note that this notion of justification applies to *methods/criteria*. A method is justified simply if using that method *in fact* leads to good (or at least better) theories, theories that deliver understanding. This notion of justification is distinct from, though related to, the notion of justification that applies to *beliefs*. We will see more of this in Chapter Two.

[18] For more on the methods of evaluating theories in science, see Braithwaite (1953), Kuhn (1962), Popper (1959), Quine (1960), Van Fraassen (1980), and Woodward (2014).

satisfying, we will want to know why our practice is justified. By digging deep into our practice of evaluating theories and explaining why that practice is justified, we can develop a theory of what makes a theory good.

Such a theory is a tall order. It is certainly not my ambition to give a complete accounting of all justified methods and criteria of choosing better theories. But it is my ambition to make progress on this score by analyzing the two candidate theoretical virtues I gestured at above: one candidate method of choosing better theories and one candidate criterion for choosing better theories. The method of choosing better theories that I consider is reflective equilibrium (which is a method that expands upon the criterion of "fit with the data") and the criterion of choosing better theories that I consider is simplicity. (For ease of expression, I henceforth refer to "methods of theory evaluation" when I actually mean "methods or criteria of theory evaluation". This is a harmless elision, since criteria are a subspecies of methods, i.e., all criteria are methods, even if not all methods are criteria). In giving an account of these two methods of theory evaluation I reach a surprising conclusion. That is, I discover that the methods we use are predicated upon good theories having a surprising feature. I conclude, namely, that both of these methods require the use of aesthetic judgment—judgments of beauty—if they are to be complete methods of theory evaluation. That is, I conclude that in order for these methods to genuinely identify better theories, they require that the person deploying those methods use aesthetic judgment to evaluate the candidate theories. This means that the beauty of a theory can be part of what makes that theory better (at least in the case of these two methods). At first, this might seem harmless. Beauty is plausibly good, so any object that possesses the property beauty is plausibly thus better, all else being equal. But this was not the kind of thing we had in mind when we asked "What makes a theory better or worse?" We are not especially interested in any property that makes a theory better in any way; we are interested in what makes a theory theoretically better or worse. That is, we want to know what makes a theory better or worse at being a theory, better or worse at fulfilling its purpose as a theory. To find that the beauty of a theory can be part of what makes that theory theoretically better would be surprising.

Indeed, the beauty of a theory might seem like a red herring to the theoretical goodness of a theory and this might be used to conclude that any method that requires aesthetic judgment is thereby an unjustified method of theory evaluation.   So, if I conclude that reflective equilibrium and simplicity require the use of aesthetic judgment in order to be complete methods of theory evaluation, then I also need to argue that the use of aesthetic judgment does not ipso facto render a method of theory evaluation unjustified (if reflective equilibrium and simplicity are to survive as justified methods of theory evaluation). Making such arguments and reaching the conclusion that the beauty of a theory can be part of what makes that theory theoretically better (at least in the case of the two methods under consideration) is the goal of this dissertation.

Section 3: Outline of Chapters

Thus, to sum up and to give a road map for this dissertation, the purpose of this dissertation is to develop a partial theory of theories, with particular attention to the last piece of a theory of theories: an account of what makes theories (theoretically) better or worse. To make progress on an account of what makes theories better or worse, I examine two candidate methods of theory evaluation. In Chapter Two, I consider the going accounts of the method of reflective equilibrium. These accounts purport to tell us how the method works and why it is a justified guide to better theories. In Chapter Three, likewise, I consider the going accounts of the criterion of simplicity. Again, these accounts purport to tell us how the method works and why it is a justified guide to better theories.

Both of these methods claim to be justified guides to arriving at better theories. If that claim is correct, which we can assess by examining the accounts of the method in question, then these methods give us a window into what makes a theory better. In my examination of these methods, however, I argue that a lacuna remains in each of the going accounts of these methods. These lacunae mean that the accounts of the methods are incomplete, and that the methods cannot yet claim to be justified methods for choosing better theories. In Chapter Four, I argue that both of these lacunae can be filled with the same thing: judgments of beauty. By building the use of judgments of beauty into both the methods of reflective equilibrium and of simplicity, I argue that we can have complete accounts of how those methods work and why they are guides to better theories. But this conclusion raises its own difficulties: The use of judgments of beauty might seem to automatically disqualify any candidate method for choosing better theories. If that were so, then my strategy for filling the lacunae with judgments of beauty would be a solution just as bad as the problem. We still would not have a satisfactory account of the candidate methods and so still would not have arrived at justified methods of choosing better theories. This would mean that we would have made no positive progress on our initial question: "What makes a theory better or worse?" We would only have found that two candidate methods are, in the end, not justified methods of choosing better theories.

Thus, in Chapter Five, I argue that the use of judgments of beauty does not automatically disqualify any candidate method for choosing better theories. That is, I argue that the beauty of a theory can justifiably be part of how one evaluates the goodness of that theory, at least when used as part of the two candidate methods under consideration. I argue for this using a kind of transcendental argument: because judgments of beauty are a necessary part of our evaluation of theories, the use of judgments of beauty is a justified part of our choosing better theories. Given all this, I conclude that philosophers use judgments of beauty when they use reflective equilibrium or simplicity to evaluate

theories and that this use of aesthetic judgment is a justified means of evaluating theories.

CHAPTER TWO
REFLECTIVE EQUILIBRIUM

Section 1: Introducing Reflective Equilibrium

To make progress on an account of what makes theories better (and thus to develop a partial theory of theories) we can examine particular methods of theory evaluation, since using these methods is meant to get us to better theories. Reflective equilibrium is a powerful and popular method of theory evaluation, particularly for theory construction. The method dictates, roughly, that one should work back-and-forth between one's common sense judgments (technically called, in the context of reflective equilibrium, "considered judgments"), on the one hand, and powerful and plausible principles, on the other hand. One works back and forth between these, revising where necessary, until one arrives at an equilibrium, a coherence between one's common sense judgments and the principles. The coherence that constitutes an equilibrium isn't mere logical coherence, but some kind of richer fit between the common sense judgments and the principles. The principles that are arrived at in such an equilibrium form the backbone of a theory. Such a theory is what the method identifies as a good theory. Like most methods of theory evaluation, there is no guarantee that one unique theory will be picked out by using the method; a variety of equilibriums could be achieved. The method is not meant to be an algorithm, but will require judgment about which theory can sustain the best reflective equilibrium. Other methods of theory evaluation, of course, can be deployed to help in the final judgment.

The method of reflective equilibrium is the analogue of the scientific method: Using the method means building theories with an eye to fitting all the relevant facts together. For the scientific method, the facts are data: observations and the like. For reflective equilibrium, the facts are our considered judgments (more on those below). From this we can see that the method is particularly useful when it comes to building theories. We can gather all the relevant facts and then proceed to build a theory that brings all those facts together. We can also see that this method is a method worked up from the analogue of the criterion of "fit with the data". The "data" for philosophical theories is of course quite different from the data of scientific theories. Nevertheless, "fit with our considered judgments" functions as the same kind of criterial constraint for philosophical theory building as does "fit with the data" for scientific theory building. Reflective equilibrium is the method built around this constraint. Although it is particularly suited for theory building, the method can also be used to select between competing theories: competing theories can

exhibit greater or lesser degrees of coherence with our considered judgments. Thus, as I discussed in the previous chapter, I continue to refer to reflective equilibrium as a general method of theory evaluation.

John Rawls coined the term "reflective equilibrium" and developed the method; it is found as early as his dissertation work.[19] But Rawls does not simply invent the method from thin air: he explicitly identifies the method in the work of Nelson Goodman.[20] In that context, Rawls also notes that the method is not limited to moral philosophy: Goodman uses it to evaluate competing theories of deductive and inductive logic. What's more, we can retrospectively identify usage of the method throughout the history of philosophy. Rawls suggests it goes as far back as Aristotle.[21] The breadth of application of the method is observable in contemporary philosophy. Although it is most often invoked in social and political philosophy, it finds natural extension from that domain into moral philosophy. And one can even find it explicitly invoked in utterly separate domains of philosophy. For instance, John Martin Fischer and Mark Ravizza explicitly deploy it has their method of choice in developing their now famous account of control in the philosophy of action.[22] Although reflective equilibrium is not without its controversies, as we will see, it enjoys widespread use and endorsement. Indeed, if Rawls' historical claim is to be believed, it has done so for quite some time.

In discussions of reflective equilibrium it is easy to slide into Rawls exegesis—his use of the method and its place in his project—given that he is responsible for explicitly developing the method and for bringing it to contemporary philosophy's attention. But if we take Rawls' own comments seriously—that the method transcends his project, moral philosophy, and contemporary philosophy—then we should be able to directly examine the merits and demerits of the method. Such a discussion has emerged in recent literature. Although the discussion often takes place in Rawls' shadow and limits itself to application in moral philosophy, it is an attempt to directly spell out how the method works and whether it is a justified guide to better theories. Norman Daniels has emerged as the prime proponent of reflective equilibrium. He has attempted to defend it from a variety of critics, perhaps chief among them Richard Brandt. Although Daniels' treatment of reflective equilibrium is promising, in many ways it neglects core features and motivations of reflective equilibrium as it features in Rawls' project. Anthony Laden has offered another interesting and powerful reconstruction, which does more justice to the specific features of reflective equilibrium critical to Rawls' project.[23] But Daniels' account of the method itself (as opposed to his account of how the method features in

---

[19] See Mandle and Reidy (2013).
[20] Rawls (1971), p. 18. See Goodman (1955), pp. 65-68.
[21] Rawls (1971,) p. 45n.
[22] Fischer and Ravizza (1998), pp. 10-11. See Sehon (2005) for another example unrelated to moral philosophy.
[23] See Laden (2013).

Rawls' project) is compatible with Laden's reconstruction. It is beyond the scope of this project to defend that claim; I must leave it for another time. My chief purpose here is not to reconstruct reflective equilibrium as it features in Rawls' project, which would do well to begin with Laden's reconstruction. Rather, my purpose is to evaluate the structure and prospects of reflective equilibrium as a method of theory evaluation as it can be pulled apart from Rawls' project. That task does well to draw on Daniels' reconstruction and defense of reflective equilibrium.

Section 2: Daniels' Reconstruction of Reflective Equilibrium

Daniels has reconstructed reflected equilibrium in a series of articles that are brought together in a synoptic monograph.[24] Drawing on all of these gives a complete picture of Daniels' account of reflective equilibrium and why he thinks it is a justified guide to better theories. Daniels' main concern in his work is to defend reflective equilibrium as a general method of theory evaluation. His interest in reflective equilibrium thus largely overlaps with mine and so I draw heavily on his work. But this should not be taken to indicate wholesale endorsement of his account of reflective equilibrium, nor that I mean to be merely reconstructing his account of the method. I analyze the method in general terms and merely draw on several of Daniels' helpful discussions.

Daniels' summary statement of the method, which is a useful general statement of the method, is as follows:

> The method of wide reflective equilibrium is an attempt to produce coherence in an ordered triple of sets of beliefs held by a particular person, namely, (a) a set of considered moral judgments, (b) a set of moral principles, and (c) a set of relevant background theories.[25]

In someone who has achieved wide reflective equilibrium, each of these types of beliefs will cohere with the other two types. And each group of beliefs is revisable in light of considerations generated from another group. The goal of the method is to work back and forth between these beliefs, connecting them and revising them, until a coherence or "fit" between the beliefs is achieved.

---

[24] See Daniels (1979), (1980a), (1980b), (1983), and (1996).
[25] Daniels (1979), p. 258.

Section 3: Clarifying the Method

There are three features of reflective equilibrium that should be clarified. First, there are several types of reflective equilibrium in Rawls' and subsequent work. We saw this already in Daniels' formal account above: he offers an account of wide reflective equilibrium. This is distinguished from narrow reflective equilibrium. The distinction is from Rawls: He says that narrow reflective equilibrium is the equilibrium which settles on the theory of justice (presumably we could generalize to other phenomenon besides justice; see the next clarification) that requires fewest revisions to one's considered judgments. Such narrow reflective equilibrium, however, neglects to compare how alternative theories of justice might fare in a reflective equilibrium and does not consider the force of various arguments for those theories of justice. Using Daniels' terminology: narrow reflective equilibrium is coherence only between (a) and (b), that is, between considered moral judgments and moral principles, leaving aside background theories. We can thus think of Daniels' background theories as providing various arguments for various theories of the phenomenon under consideration.

It is also worth noting that Rawls distinguishes between two other kinds of reflective equilibrium. Reflective equilibrium as I have discussed it so far, wide or narrow, is a method or project for one person. It is an attempt to find an equilibrium between one's own considered judgments and plausible principles. But Rawls is also interested in reflective equilibrium between persons. The first kind of reflective equilibrium between persons is general reflective equilibrium. It is achieved when all members of a society, in each of their own reflective equilibriums, affirm the same theory of justice. Combine general reflective equilibrium with the stipulation that everyone is also in their own individual wide reflective equilibrium and we get full reflective equilibrium. When full reflective equilibrium is achieved, "not only is there a public point of view from which all citizens can adjudicate their claims, but also this point of view is mutually recognized as affirmed by them all in full reflective equilibrium".[26]

So, narrow and wide reflective equilibrium obtain within a given individual, while general and full reflective equilibrium obtain among a group of individuals. These two latter kinds of reflective equilibrium—between all citizens of a society—are important to Rawls and his particular project. Given some of the special problems of justice, it is easy to see why: when we are choosing a theory of justice—which will authorize certain kinds of control over peoples' lives—it is important that it be a theory of justice that everyone can understand, see themselves in, and remain committed to the system that results. Constructivists about morality will make analogous moves about the role of morality. So they, too, may pursue full or general reflective equilibrium in using the method to arrive at a better theory of morality. Whether one sees narrow/wide or full/general (that

---

[26] Rawls (2001), p. 31.

is, intrapersonal or interpersonal) reflective equilibrium as the more important form of the method will largely turn on the nature of the phenomenon one is theorizing about.

My purpose is to examine reflective equilibrium as a general method of theory evaluation, not to advance any particular understanding of it. Thus, my analysis is agnostic about which form of reflective equilibrium—narrow/wide or full/general—is the preferable or more important form of the method. I will continue to use the unqualified term "reflective equilibrium" throughout. The analysis can be adjusted to whatever form of reflective equilibrium is more appropriate for a given case.

Second, as I noted at the outset, reflective equilibrium is not limited to the political or moral domain. Although Daniels does use the term "moral" in his formal account of reflective equilibrium above, this reflects the more specific purpose of that particular paper, which is to follow up on Rawls' work and offer reflective equilibrium as a promising method for ethics. But he is also clear in this earliest piece, and even more so in later work, that the method is not limited to ethics. He explicitly compares it to a method of scientific theory evaluation:

> [In reflective equilibrium,] as in science, judgments about the plausibility and acceptability of various claims are the complex result of the whole system of interconnected theories already found acceptable. My guess—I cannot undertake to confirm it here—is that the type of coherence constraint that operates in the moral and nonmoral cases functions to produce many similarities: we should find methodological conservatism in both; we will find that "simplicity" judgments in both really depend on determining how little we have to change in the interconnected background theories already accepted (not on more formal measures of simplicity); and we will find in both that apparently "intuitive" judgments about how "interesting", "important", and "relevant" puzzles or facts are, are really guided by underlying theory.[27]

In another of his articles, Daniels also develops an analogy between reflective equilibrium and a method used in logic to choose between competing alternative logics.[28] And in his later work, Daniels makes it very clear that he aims to articulate and defend a method that can be applied in a variety of domains.[29] This is in the spirit of Rawls, given that he says that the method transcends his particular project, and in keeping with the fact that reflective equilibrium has enjoyed uptake in other domains in philosophy. I follow Daniels in this respect and examine reflective equilibrium as a method of theory evaluation that is neutral with respect to philosophical domain.

---

[27] Daniels (1979), p. 262.
[28] Daniels (1980a).
[29] Daniels (1993).

Third, and more substantively, it is important to clarify the three types of beliefs that play a part in reflective equilibrium. The first type of beliefs is "considered judgments". These judgments are similar to (although not identical to) what are often called "intuitions" in philosophy: common sense judgments about ordinary cases, about general rules, and about theories.[30] Rawls says that considered judgments are those judgments that we make "intuitively" and "with the greatest confidence". He defines them: "Considered judgments are simply those rendered under conditions favorable to the exercise of the sense of justice, and therefore in circumstances where the more common excuses and explanations for making a mistake do not obtain…The criteria that identify these judgments are not arbitrary. They are, in fact, similar to those that single out considered judgments of any kind".[31] He gives as an example of considered judgments: "We are confident that religious intolerance and racial discrimination are unjust".[32] Such considered judgments are "provisional fixed points which we presume any conception of justice must fit".[33] Although all these examples of considered judgments are from the moral domain, we have considered judgments in many domains.

Although it is natural to compare considered judgments with the more common notion of intuitions, there are some important differences. Unlike some accounts of intuition, considered judgments are not self-evident. That is, they do not have inviolable justificatory status. Rather, they stand before the bar of critical reason like any other piece of a theory. Although they are not meant to be self-evident, it is clear that there is something justificatorily special about considered judgments: Different accounts of reflective equilibrium hold that considered judgments either enjoy some justificatory status or do not need any justificatory status. The justificatory status of considered judgments—whatever it is—is important for reflective equilibrium and the subject of controversy. We'll turn to that shortly in Section 4. There is another important difference between considered judgments and many accounts of intuitions. Considered judgments are not verdicts about fantastical cases, such as trolley cases. Such verdicts are often called intuitions, but considered judgments are meant to be spontaneous, stable upon reflection, and tied to regularly experienced human conditions. Clearly, verdicts about trolley cases and the like fail to meet this last condition. Considered judgments are meant to manifest our freely expressed human nature. Given that verdicts about trolley cases and the like are not tied to regularly experienced human conditions, they do not contribute to the project of achieving understanding by means of reflective equilibrium.

---

[30] For discussions of the role of intuitions or considered judgments in theorizing (particularly moral theorizing), see Brandy (1979), Cuneo (2011), DePaul and Ramsey (1998), Elgin (1993), Rawls (1975), Sencerz (1986), Sidgwick (1907), and Singer (2005).
[31] Rawls (1971), p. 42.
[32] Rawls (1971), p. 17.
[33] Rawls (1971), p. 18.

The second type of beliefs in reflective equilibrium is "principles". Principles are generalizations that explain the target phenomenon and can yield a somewhat determinate judgment about that phenomenon. Rawls' two principles of justice are, of course, the key principles in his use of reflective equilibrium. Other examples of principles put forward in the moral domain include Kant's categorical imperative, Aristotle's rule of the golden mean, and the golden rule. These principles purport to explain morality and yield somewhat determinate judgments about what is or is not the right thing to do in a given circumstance. Of course, principles are not only important parts of ethical explanations: They are invoked in virtually any explanatory system. Recall how the principles of statistics and human behavior formed the backbone of the theory about who killed Mr. Boddy. Principles are core distillations of a system that are meant to explain that system and adjudicate over a wide range of particular cases within that system. Just as we have considered judgments in many domains, we have principles in those various domains that are meant to explicate those judgments. The method of reflective equilibrium dictates that principles should be constructed and selected in a way that coheres with and explains our considered judgments. Principles are to be revised in light of clashes with considered judgments, but so too should considered judgments be revised in light of powerful and plausible principles.

Principles can be arrived at in a variety of ways. Most straightforwardly, we can have considered judgments about which principles are plausible and which are not. In the same way that we arrive at considered judgments about particular cases, we can have judgments about the plausibility of principles that are spontaneous, stable upon reflection, and tied to regularly experienced human conditions. We can arrive at principles in other mundane ways: we can draw them out of the history of philosophy or invent them through sheer innovation. Or, more ambitiously, principles can be delivered by more complicated construction procedures. Rawls, for instance, seems to deploy some kind of construction procedure in his theoretical system to arrive at his favored two principles of justice. (Although the exact nature and role of this construction procedure in Rawls' project is the subject of debate.)[34] Any and all of these methods for constructing principles can be used in building a theory. Reflective equilibrium says that the candidate principles of a theory, however they are arrived at, must cohere with our considered judgments and relevant background theories. The principles of a theory arrived at through reflective equilibrium are to be evaluated not by their etiology but by their fit in the system.

The third type of beliefs in reflective equilibrium is "background theories". Background theories are relatively complete, free-standing theories of other phenomena, which are related to the target phenomenon of a reflective equilibrium. Just as we have considered judgments in many domains, and principles in those domains explicating those judgments, we also often have background theories in those domains generating those principles. For example:

---

[34] See Laden (2013).

Rawls argues that a reflective equilibrium about the principles of justice would settle on his favored two principles of justice (at least compared to the going alternatives). Several background theories feature in this argument, which Daniels highlights: a theory of procedural justice, a theory of the person, and a theory of the role of morality in society, with the notion of a well-ordered society falling out of the latter two.[35] If Daniels' interpretation of Rawls is correct, these theories are developed independently (or at least partially independently), and then feature in the reflective equilibrium that settles on the two principles of justice.[36] Of course, the purpose of deploying reflective equilibrium is to arrive at better theories, so there might be something confusing about using theories as part of the method. But this involves nothing more than a nested structure of theories: some background theories (we might also call them "sub-theories") feature in the reflective equilibrium that delivers a larger theory. As an example, we can again look to explaining the death of Mr. Boddy: The principles of human behavior are probably better described as coming together to form a rudimentary theory of human behavior: an explanation of the ends that humans typically pursue and the means that they typically take to those ends. This theory of human behavior is then used as a sub-theory or background theory in evaluating our theory of who killed Mr. Boddy. Thus, reflective equilibrium involves looking for coherence and fit not only between considered judgments and principles, but also between considered judgments and relevant background theories as well as principles and relevant background theories. Each can be revised in light the others. The content of relevant background theories depends, of course, on the target phenomenon that one is trying to hold in reflective equilibrium.

We can now give a complete and rudimentary example of how reflective equilibrium is used. Begin with an object of inquiry: Who killed Mr. Boddy? We want an answer to that question and an explanation of that answer. The answer will thus come in the form of a theory: a central claim about who did it and a system of supporting claims about how and why they did it. If we are using reflective equilibrium to answer this question, then we come to a candidate theory by reflectively working among our considered judgments, plausible principles, and background theories. In this case, the analogue of our considered judgments are the facts about the case: Where Mr. Boddy was found, how he died, who was nearby, etc. The relevant plausible principles are, among others, the principles of statistics: Mr. Boddy could have been struck by lightning is such a way as to produce the state he was found in, but this was extremely

[35] Daniels explores how these background theories feature in Rawls' arguments in Daniels (1980b).
[36] Daniels is also concerned to argue that reflective equilibrium requires the addition of an "independence constraint", which says that the considered judgments that support target principles must be separate from the considered judgments that support relevant background theories. If this constraint is not respected, then the reflective equilibrium risks viciously circularity. We'll come to the question of the potential circularity of reflective equilibrium in Section 4.

statistically unlikely. The relevant background theories are, among others, the basic theories of human behavior: Wealth is an end that many people have and people have been known to kill if it is a means to that end. Suppose that we arrive at the theory that Colonel Mustard did it in such-and-such a way and for such-and-such a reason after reflectively working among these considerations, seeking to make them all cohere. The method of reflective equilibrium says that if that theory coheres with all relevant considered judgments, plausible principles, and background theories, then that theory is a good theory.

Section 4: Is Reflective Equilibrium a Justified Method?

Recall that I said in the previous chapter that the purpose of theories is to explain a target phenomenon so that we can understand that phenomenon. If a theory is the result of using a justified method of theory evaluation, then that theory is a good theory (or at least better than its rivals). That is, that theory is a source of explanation and understanding. When we examine an account of a particular method of theory evaluation, we are looking to see whether and how it is a justified method, whether and how using that method can lead to theories that deliver explanation and understanding.[37] The literature on reflective equilibrium can be framed in this way: it is an examination of whether and how using reflective equilibrium can lead to theories that deliver explanation and understanding. But the literature on reflective equilibrium explicitly frames the question of whether reflective equilibrium is a justified method in terms of whether using the method leads to justified beliefs. The beliefs that could be justified by using reflective equilibrium are the beliefs that compose a theory that is marked as better by using reflective equilibrium. In the case of reflective equilibrium, such beliefs that compose a theory are considered judgments, principles, and background theories. On this way of framing things, whether reflective equilibrium is a justified method depends on whether a theory's being picked out by using reflective equilibrium justifies the beliefs that compose that theory. Obviously this is a different way of framing the purpose of methods of theory evaluation compared to the account given above. But it is not a stretch to think that a theory that delivered explanation and understanding would also justify the beliefs that compose that theory and hence that a method that led to the one would also lead to the other. For most of this dissertation, I continue to frame the purpose of methods of theory evaluation in terms of explanation and understanding. But, in the interest of accurately representing the literature on reflective equilibrium, I frame the discussion of the purpose of reflective

---

[37] Remember that this notion of a *justified method* is distinct from the notion of a *justified belief*. A justified method is simply one whose use results in theories that deliver explanation and understanding.

equilibrium in terms of arriving at theories that provide justification for the beliefs that compose those theories.

So, if reflective equilibrium is to be a justified method (in this latter way of framing things), a theory's being the result of using reflective equilibrium must justify the beliefs that compose that theory. How could a theory's being the result of using reflective equilibrium justify the beliefs that compose that theory? Rawls and Daniels devote considerable attention to this question. They argue that the justification that reflective equilibrium aims at is not foundationalist, but has elements both of coherentist and common sense accounts of justification.[38]

Both Rawls and Daniels argue that reflective equilibrium does not rely on a foundationalist form of justification.[39] A foundationalist justification is one that begins from some epistemically privileged starting point. Justification is then transferred to other beliefs by being derived from that starting point. Rawls makes it clear that reflective equilibrium does not involve this kind of justification. He says, "[The justification of a conception of justice] is a matter of the mutual support of many considerations, of everything fitting together into one coherent view".[40] He later expands on this by distinguishing two traditional views of justification.[41] The first he calls Cartesian: it relies on self-evident foundational premises from which all subsequent moral theorizing derives its justification. The second he calls naturalism: it introduces definitions of moral concepts in non-moral terms and then moral theorizing derives its justification from this non-moral foundation. Thus, both of these methods are foundationalist: the one thinks that moral self-evident truths constitute the foundation of moral theory, the other thinks that non-moral (and hence justifiable by ordinary means) truths constitute the foundation of moral theory. Rawls says that he has not adopted either of these conceptions of justification. There is no claim that the basis of his theory is self-evident or non-moral: they are contingently held and yet thoroughly moral. He says: "Therefore we do better, I think, to regard a moral theory just as any other theory, making due allowances for its Socratic aspects".[42] He expands on this comment in a footnote: He cites Quine's view of how theories are justified, which is applied to ethics by Morton White.[43] Quine's view that he seems to have in mind is the Quine/Duhem thesis: the view that when we evaluate a scientific

---

[38] By a "common sense" account of justification, I simply mean the view that accords *some* justificatory status to considered judgments. This is not meant to be the view that accords *indefeasible* or *self-evident* justificatory status to intuitions. As I distinguished earlier, considered judgments are not intuitions in this sense.

[39] This literature is obviously closely related to the literature in epistemology on whether justification is foundationalist or coherentist. See Sosa (1980) for an introduction to this debate. See also DePaul (1986) for another argument that reflective equilibrium needn't be foundationalist.

[40] Rawls (1971), p. 19.

[41] Rawls (1971), §89.

[42] Rawls (1971), p. 506-7.

[43] See White (1956).

24

theory in light of a competing observation, we are always free to change any part of the theory to accommodate the observation, including our interpretation of the observation itself.[44] That is, we don't necessarily need to change the exact part of the theory that the observation conflicts with, or the part of the theory that is under scrutiny. This seems to be manifest in Rawls' project as the thought that no part of the theory—including our most treasured considered judgments—are immune from revision. In principle, any part of the theory can be revised to resolve a tension. White explicitly takes up this view of justification and argues that it should be used in ethics as well as science. White says:

> A radical epistemological difference of kind [between ethical, scientific, and logical principles] can not be drawn. According to this view, every individual comes to his moral problems with a complex of moral and scientific beliefs that are under test or scrutiny, a scientific theory and moral code that lead him by logic to say what will be and what ought to be….There is no doubt that the principles of logic are less frequently challenged than are those of ethics, and I suspect that those of ethics are less frequently challenged with a given culture than are those of physics, though this might be doubted by some. It is not surprising, therefore, that both the principles of logic and those of ethics are said to be established by peculiar activities like "seeing" the connections between meanings of intuiting the presence of a non-natural quality. What the proposed view achieves is a way of avoiding this invention of entities and of queer relations between them that are detected in queer ways. We dispense not only with the need for platonic meanings but also with the need for non-natural qualities. We recognize the connections between logic, ethics, and physics by seeing how they function together.[45]

It is such a coherentist view of justification, with its characteristic back-and-forth revision, that Rawls seems to have in mind when describing theorizing as "Socratic".[46] White obviously ties it with a presumptive empiricist or linguistic method of theorizing, and that seems somewhat dropped in Rawls (although he does at one point draw an analogy between reflective equilibrium and the linguistic method of formulating rules of grammar, paralleling the work of Nelson Goodman).[47] Indeed, in rejecting the naturalistic method—justification proceeding from non-moral premises—Rawls does seem to reject this naturalistic

---

[44] See Quine (1960), (1969), (1975).

[45] White (1956), p. 256.

[46] Given that Socrates conceived of philosophy as the pursuit of self-understanding, this comment also might invoke the second-personal conception of justification that I discussed above. That is, the characteristic back-and-forth theorizing of reflective equilibrium can be understood in the spirit of the Socratic pursuit of self-understanding or expanded to be pursuit of mutual-understanding.

[47] Rawls (1971), p. 41. See Goodman (1955).

reading of coherentist justification. Instead, against both the naturalistic and the Cartesian method, Rawls says:

> I have not proceeded then as if first principles, or conditions thereon, or definitions either, have special features that permit them a peculiar place in justifying a moral doctrine. They are central elements and devices of theory, but justification rests upon the entire conception and how it fits in with and organizes our considered judgments in reflective equilibrium. As we have noted before, justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view. Accepting this idea allows us to leave questions of meaning and definition aside and to get on with the task of developing a substantive theory of justice. The three parts of the exposition of this theory are intended to make a unified whole by supporting one another…[48]

The emphasis on fittingness, coherence, and unity is striking. It is clear that the method of reflective equilibrium is meant to provide justification understood in a coherentist way.

Daniels argues in an analogous way that reflective equilibrium is not foundationalist. He argues that reflective equilibrium is not strictly foundationalist, since no considered judgments (or principles or background theories, for that matter) are regarded as inviolable. All of them are revisable, in principle. Especially when pushed to wide reflective equilibrium, which makes essential reference to background theories, there is plenty of ground for revision. He says, "In seeking wide reflective equilibrium, we are constantly making plausibility judgments about which of our considered moral judgments we should revise in light of theoretical considerations at all levels".[49] We are constantly testing our considered judgments against both principles and theories, and judging their acceptability based on the system that they can coherently fit into. This is a complicated, holistic judgment, where the system is judged as a coherent whole and not where any one part has privileged status.

Reflective equilibrium's reliance on considered judgments has seemed to some, however, to invoke a foundationalist account of justification: These considered judgments are justified in some way, which then pass their justificatory status onto the principles and background theories in reflective equilibrium. This can seem directly contrary to the conclusion that reflective equilibrium only deploys a coherentist account of justification. But Daniels argues that justification for considered judgments should be distinguished from

---

[48] Rawls (1971), p. 507, emphasis added. Note that the "three parts of the exposition of this theory" refers to the largest tripartite division of Rawls' presentation of his theory; this should not be confused with Daniels' formalized three "sets" of beliefs account of reflective equilibrium.
[49] Daniels (1979), p. 267.

foundationalism.[50] Common sense justification requires conferring some justificatory status to our considered judgments, while foundationalism requires conferring privileged justificatory status to something (for instance, our considered judgments). Daniels thinks that reflective equilibrium must award some justificatory status to considered judgments—they are provisionally justified—but insists that they are not foundational. He says, "No considered moral judgments at any level are taken to be unrevisable, that is, strongly foundational; moreover, they are subject to revisionary pressures from considerations at all levels."[51] The coherentism of reflective equilibrium allows considered judgments to be brought into the justificatory system without relying on them as a foundation.

But even this non-foundationalist version of common sense justification faces a traditional objection to common sense justification. Richard Brandt raises this objection to reflective equilibrium: Why think that considered judgments are worthy of any justificatory status?[52] Brandt proceeds by way of an analogy to science. In scientific theory evaluation, we choose theories that account for the observational data. This is analogous to reflective equilibrium's choosing theories that account for our considered judgments. But the observational data of science is justified by the traditional means of observation.[53] There are no such "facts of observation" for our considered judgments. We have no reason to think that even our considered moral judgments are anything more than mere prejudice. Organizing such prejudices into a coherent system of theories and principles would not justify them. And thus, reflective equilibrium would not be a justified method of theory evaluation.

Daniels responds in two main ways. First, he argues that an analogy between observational data and considered moral judgments is strained: considered moral judgments are still theory-laden. That is, we vindicate and explain our moral judgments by reference to reasons; they are supported by a theoretical structure, not by having been formed under proper conditions (which is, perhaps, how we support observations).[54] Considered moral judgments call out for explanations and we are accustomed to raising theoretical considerations in support of them. Second, Daniels argues that it is precisely wide reflective equilibrium about a moral theory that could provide the apparatus for distinguishing credible considered moral judgments from non-credible ones. That is, an answer to the credibility question about considered moral judgments

---

[50] Daniels uses the term "intuitionism", but for reasons highlighted earlier, I use the term "common sense justification".

[51] Daniels (1980b), p. 102.

[52] See Brandt (1979). See also Sencerz (1986) for a version of this objection.

[53] This is already too hasty; many "observations" in contemporary physics are theory-laden interpretations of data that only indirectly speaks to the target phenomenon. This is consonant with Daniels' reply to Brandt's objection, examined below.

[54] Although Rawls, perhaps, sometimes pushed for this analogy. See Rawls (1971), p. 42.

may be forthcoming, and it may be precisely wide reflective equilibrium that delivers them. He further suggests that even observational data are only distinguishable as credible or not by reference to a well-established body of theory, something like wide reflective equilibrium in a particular field of science. This would go further than the previous point, and make observations themselves also more theory-laden. Daniels concludes that reflective equilibrium can stand up to this "no credibility" objection about considered judgments, given that the coherence constraint inherent in wide reflective equilibrium acts as credibility check on considered judgments. Indeed, he thinks that wide reflective equilibrium constitutes a method very similar to scientific method, which is perhaps to turn Brandt's objection back on itself. He says:

> My reply to the 'no credibility' criticism points again to a strong similarity in the way coherence constraints on theory acceptance (or justification) operate in the two domains [moral theory and natural science], despite the disanalogy between observation reports and considered moral judgments…In rejecting the view that wide reflective equilibrium merely systematizes a determinate set of moral judgments, and arguing instead for the revisability of these inputs, I suggest that wide equilibrium closely resembles scientific practice. Neither in science nor in ethics do we merely 'test' our theories against a pre-determined, relatively fixed body of data. Rather, we continually reassess and reevaluate both the plausibility and the relevance of these data against theories we are inclined to accept.[55]

Thus, the coherentism of reflective equilibrium deflects this traditional objection to common sense justification. By enmeshing our considered judgments in a larger coherence of principles and theories, we gain the critical leverage needed to adjudicate their credibility.

But deploying the coherentism of reflective equilibrium to rescue it from the traditional challenge to common sense justification raises a traditional objection to coherentism: Why think that having coherence among beliefs justifies them? That is, it doesn't seem that mere coherence among beliefs can confer justification, since any group of beliefs (even repugnant systems of beliefs like Nazism) could form a coherent system. Daniels responds to this objection by arguing that we should be using wide reflective equilibrium, which renders the circularity of coherentism non-vicious. Wide reflective equilibrium is distinctive in that it introduces a third type of beliefs to be held in coherent equilibrium: background theories. This already makes some progress: the match between principles and considered judgments is not simply a circle between the two. It also involves background theories. But perhaps this is just a wider circle, but no less vicious. It is at this point that Daniels develops an independence constraint

---

[55] Daniels (1979), p. 273.

that he argues should be conjoined with reflective equilibrium.[56] This independence constraint requires, roughly, that the group of considered judgments that supports a set of principles in a reflective equilibrium be distinct from the group of considered judgments that supports a set of background theories. That is, the independence constraint requires that there be at least two distinct groups of considered judgments in any wide reflective equilibrium: one that supports the principles and one that support the background theories. Thus, there are multiple points of input for our considered judgments, which makes the interconnections that constitute a coherence in reflective equilibrium more complicated than a simple circle. Daniels suggests, plausibly, that this renders the coherentism of reflective equilibrium non-vicious. We can gloss this independence constraint as a reemphasis of the common sense justification of reflective equilibrium: reflective equilibrium is meant to proceed from our considered judgments. It is this common sense justification that grounds the system of coherence that is meant to be developed in reflective equilibrium: our considered judgments check the coherence being developed in reflective equilibrium at multiple points. If only coherence is required, then systems of justification can get off the ground in complete isolation from our considered judgments, without any justificatory traction. The invocation of common sense justification, however, gives reflective equilibrium that necessary traction. It requires that the coherence developed in reflective equilibrium include our considered judgments.

Thus, both common sense justification—partially relying on considered judgments—and coherentist justification—evaluating those considered judgments as part of a coherent whole—are important to the justificatory success of reflective equilibrium. The combination of coherentist and common sense justifications that constitutes reflective equilibrium allows it to defend against both types of traditional worries: coherentism blocks the potential prejudice of common sense justification by making all parts of the system potentially revisable, and common sense justification blocks the potential vicious circularity of coherentism by grounding the system in considered judgments. It is a reflective back-and-forth among our considered judgments—with the aim of settling on a coherent equilibrium—that delivers the justificatory power of reflective equilibrium. That is, this coherentist/common sense structure of reflective equilibrium means that a theory's being the result of using reflective equilibrium justifies the beliefs that compose that theory. This means that reflective equilibrium is a justified method of theory evaluation.

---

[56] Daniels develops and defends the independence constraint across his work, but mostly in Daniels (1980b).

Section 5: What Kind of Coherence is Required by Reflective Equilibrium?

We have been building an account of reflective equilibrium. We examined how the method works and why it is a justified method for evaluating theories. But there is one more feature of reflective equilibrium that we must examine if we are to have a complete account of the method: coherence. The method instructs us to seek coherence between three types of beliefs: considered judgments, principles, and background theories. What kind of coherence does the method aim at? What relationship between the three types of beliefs does the method aim at? To put the question another way: What generates the pressure to revise either considered judgments, principles, or background theories? If the method merely instructed us to conjoin three groups of beliefs, it would not generate any pressure to revise them. What notion of coherence could generate pressure to revise the beliefs so they could form one coherent system?

The first natural candidate is simply logical consistency. Rawls and Daniels both speak of reflective equilibrium being "underdetermined" (channeling the Quinean genealogy of contemporary reflective equilibrium[57]). That is, they both maintain that reflective equilibrium only picks out a range of better theories; it does not determine a single uniquely good theory. Mere logical consistency obviously has this feature: it only delimits the range to exclude theories that are logically inconsistent. Nevertheless, this understanding of reflective equilibrium's coherence constraint is obviously too weak. We would hardly need such an elaborate method as reflective equilibrium to tell us merely that logically inconsistent theories aren't good theories. Nor would the range of better theories be delimited in any substantive way with such a constraint. Nor could mere logically consistency generate enough of the pressure toward revision that the method requires. Daniels explicitly says all this, and compares the coherence of reflective equilibrium to the more substantive coherence constraints found in scientific methods. He says: "Coherence involves more than mere logical consistency. As in the sciences, for example, we often rely on inference to the best explanation and arguments about plausibility and simplicity to support some of our beliefs in light of others".[58]

So, some stronger notion of coherence is needed.[59] But Rawls and Daniels do not have an account of what coherence is meant to be. In the above

---

[57] For more on theory underdetermination and the Quinean roots of reflective equilibrium, see Bergstrom (1984); Boyd (1973); Earman (1993); Glymour (1971); Laudan and Leplin (1991); Quine (1960), (1969), and (1975); and Stanford (2013).

[58] Daniels (1992), p. 2.

[59] One might think that the literature in epistemology on coherentist justification would be helpful for identifying the exact nature of coherence. See Sosa (1980). And see Swanton (1992) to see the coherentist view of justification in action. But this literature is mainly concerned with figuring out how coherentist structures—whatever they are—can deliver justification and what exactly this justification is like. It is less interested in

passage and elsewhere Daniels compares the coherence constraint to other kinds of constraints: inference to the best explanation, plausibility, and simplicity.[60] At another point he describes coherence as "economical systematicity".[61] He most commonly glosses the coherence constraint of reflective equilibrium as requiring a "fit" between the three types of beliefs. In a similar way, Rawls says: "[The justification of a theory of justice] is a matter of the mutual support of many considerations, of everything fitting together into one coherent view."[62] Again, there is a reference to everything "fitting together", but it remains unclear exactly what that amounts to. Talk of "fittingness" does seem to be a stab at answering the question "What generates the pressure to revise various beliefs?". The method pressures us to revise considered judgments, principles, and background theories until they all fit together. But this gloss does not really resolve the question, but only raises it again: What kind of "fit" between beliefs does reflective equilibrium prescribe? How exactly does fittingness generate pressures to revise? It is here that exegesis of Rawls and Daniels gives out as the best way to understand reflective equilibrium. They seem to have no answer to the related questions: "What exactly does the coherence constraint of reflective equilibrium require? What kind of coherence is prescribed for better theories? How does coherence generate pressure to revise parts of a theory?" Indeed, it is hard to find them even taking up such questions. I therefore depart from exegesis at this point and in the coming chapters propose my own positive account of how best to understand the coherence requirement of reflective equilibrium.

Remember that my main purpose is to develop a (partial) theory of theories, with particular attention to what makes theories (theoretically) better. By analyzing some specific methods for evaluating theories, we can learn something more general about what makes theories good. In this chapter I analyzed reflective equilibrium: we have seen how the method is meant to work, the various moving pieces that compose it, and why it is a justified method of theory evaluation. (The argument that it is a justified method of theory evaluation was framed in terms of justifying the beliefs that compose the theories that are arrived at by using the method. But it could be reframed in terms of delivering explanation and understanding.) But this analysis has run up incomplete at the end: What is the critical notion of coherence invoked by the method? This lacuna is the set-up for the conclusion about evaluating theories that I argue for. I will argue that this lacuna can only be filled by an aesthetic understanding of coherence. This gets us to the conclusion that reflective equilibrium is a partially aesthetic criterion of theory evaluation. This would mean that at least one of our

---

identifying precisely *what* coherence is or precisely *what* judgments of coherence are like.

[60] "Inference to the best explanation" is its own vexed criterion of theory selection. For an overview of inference to the best explanation, see Douven (2011).
[61] Daniels (1980b), p. 22.
[62] Rawls (1979), p. 19.

best methods for evaluating theories is partially aesthetic. This, in turn, would mean that part of what makes theories better is their beauty.

That is where I will go, but I am getting ahead of myself. These arguments will come in later chapters. For now, I continue to set up these coming arguments by analyzing another powerful and popular method of theory evaluation: simplicity. I analyze simplicity in the same way I did with reflective equilibrium. I also come to a similar conclusion: there is a remaining lacuna in the analysis of simplicity. I will then be in a position to develop in the following chapters my central argument: both these lacunas can be filled by an aesthetic understanding of each method. This will mean that both methods are partially aesthetic. But for now, on to an analysis of simplicity.

CHAPTER THREE

SIMPLICITY

Section 1: Introducing Simplicity

Recall that my purpose is to make progress on an account of what makes theories better.   My strategy for making progress on this issue is to analyze a couple key methods of theory evaluation. This is because such methods, whose use is meant to get us to better theories, will help illuminate what makes theories good. The second method (technically, it is a criterion) of theory evaluation that I analyze is simplicity. It is particularly used for theory selection. Compared to the method of reflective equilibrium, the criterion of simplicity is straightforward. It flat-footedly dictates that we are justified in believing the simpler of two otherwise comparable theories. Like reflective equilibrium, there is no guarantee that the simplicity criterion will evaluate one theory as uniquely good, uniquely simple. Nor is the criterion meant to operate in isolation from other methods: a ceteris paribus clause is critical for the simplicity criterion and other methods will do much work in establishing that other things are indeed equal.

The simplicity criterion is obviously also important for scientific theory evaluation. It is particularly useful for selecting between competing theories that are otherwise roughly equal. Imagine that one built two scientific theories using the scientific method: building a theory to fit as much as possible with all our observations. Or imagine that one built two philosophical theories using reflective equilibrium: building a theory to fit as much as possible with all our considered judgments. Now, if there were dramatic differences in simplicity between these two pairs of competing theories—one theory was substantially simpler than its competitor—we would use the simplicity criterion to select the simpler of the two. But this does not mean that simplicity is only relegated to this selection stage of theory evaluation: when we are building our theories—whether scientific or philosophical—we will use the simplicity criterion as a constraint on that theory building. Thus, as I discussed in Chapter One, I refer to simplicity as a general criterion of theory evaluation.

No one is responsible for coining the criterion of simplicity in the way that John Rawls coined reflective equilibrium—perhaps the closest would be Ockham's razor, which I briefly discuss below. But nor is the criterion as explicitly invoked for use in theory evaluation as often as reflective equilibrium. Unlike reflective equilibrium, simplicity is a pervasive background criterion of theory evaluation that most often goes without saying. This does not mean that it

has not been analyzed in the history of philosophy or in contemporary philosophy, nor does it mean that it is without its controversies.[63]

The criterion of simplicity is more often explicitly invoked and analyzed in one domain of philosophy: the philosophy of science. This is because, as I said above, simplicity is a pervasive background criterion of theory evaluation for natural science as well as philosophy, and philosophers of science are vested in understanding and justifying the criteria of theory evaluation that natural science uses. (Philosophers of science tend to be especially interested in simplicity because it is not easy to see how the criterion of simplicity squares with the rationalism or empiricism of science.[64]) It is not my purpose to venture too deeply into the philosophy of science; my chief concern here remains understanding methods of theory evaluation in their application to philosophy as a means to understanding what makes theories better. Nevertheless, I begin my analysis of simplicity by looking to the work of Richard Swinburne and James McAllister, both of whom analyze simplicity in the context of the philosophy of science.[65] Analyzing their accounts of simplicity will lead to my conclusion about reflective equilibrium and simplicity: that both of them need aesthetic judgment in order to be complete methods of theory evaluation. This means that at least two of our most important and powerful criteria of theory evaluation are partially aesthetic criteria. The endpoint of this chapter thus will be analogous to the endpoint of the previous chapter: the analyses of simplicity offered by Swinburne and McAllister will leave a lacuna remaining, which I argue in the next chapter is best filled by aesthetic judgment.

Section 2: Swinburne on Simplicity

Swinburne begins his analysis of the simplicity criterion by examining what makes an explanation (or a theory) better or worse.[66] He argues that there are four primary criteria that demarcate better explanations from worse. The first two criteria are a posteriori. (1) The criterion of yielding the data. For a scientific explanation, the data are observations; for a philosophical explanation, the data are something like our considered judgments as discussed in the previous chapter. The more data that the explanation implies, the better the explanation.

---

[63] For more on simplicity, see Baker (2013), Derkse (1993), Sober (1975), Swinburne (1997), and Walsh (1979).

[64] For more on simplicity as it particularly relates to science, see Bunge (1963), Chandrasekhar (1987), Feuer (1957), Forster and Sober (1994), Hillman (1962), Kemeny (1953), and McAllister (1999).

[65] See Swinburne (1997) and McAllister (1999).

[66] Swinburne uses "explanation" and "theory" interchangeably, though he mainly uses "explanation". I follow that usage in reconstructing his account, but I remain primarily focused on theories.

(2) The criterion of fit with background knowledge. The better the proposed explanation fits with our background knowledge of the relevant phenomenon, the better the explanation.[67]

Swinburne's next two criteria are a priori. (3) The criterion of content. The greater the content of the explanation, the worse it is. That is, the more that an explanation commits itself to, the deeper it is, the more vulnerable it is to being a bad explanation. (4) The criterion of simplicity. The simpler the explanation, the better the explanation. Defending this as an actual criterion of better explanation is, of course, Swinburne's task.

As always, these four criteria can pull in different directions. But Swinburne claims that if three of the four criteria are held fixed between two competing explanations, then the one that scores better on the remaining fourth criterion will be better. Of course, such idealized comparisons are rare; more commonly, success on one criterion will be traded against failure on another. This is most notable for comparisons between the content and simplicity criteria. In general, we want our explanations to be deep and not superficial. But depth is precisely what the content criterion says will make our theories worse. We can think of the simplicity criterion precisely as the counter-balancing force that can make deep theories (theories with high content which are thus more likely to be bad) justified, despite their scoring poorly on the content criterion. If we were left with only the content criteria, without the simplicity criteria, we would never be justified in holding to deeper explanations. That is, the content criteria, without the simplicity criteria to counter-balance it, would drive us to only make superficial explanations: at the limit, mere lists of observations, mere lists of considered judgments. This can be understood as part of Swinburne's argument for the simplicity criteria, which I turn to now.

Swinburne argues that if we hold the other three criteria fixed and then make comparisons between competing explanations, it is clear that we rely on the simplicity criterion to judge which explanation is better. A mathematical example makes his point clear, but it generalizes to philosophical explanations. Consider the following data:

x=1, 2, 3, 4, 5, 6
y=2, 4, 6, 8, 10, 12

The natural explanation of the data is that y=2x. But infinite other ratios also capture the data, but yield very different predictions for next data point

---

[67] Swinburne also eventually argues that this criterion—fit with background knowledge—already invokes the simplicity criterion. It is also worth noting that Swinburne's account of the criteria of explanation seems strikingly in line with the method of reflective equilibrium. Swinburne's first two criteria are reminiscent of reflective equilibrium's requirement that a theory fit with both our considered judgments and with background theories. This is another example of reflective equilibrium's pervasiveness in philosophical theorizing, even when not explicitly invoked.

(e.g., y=2x + x(x-1)(x-2)(x-3)(x-4)(x-5)(x-6)z). Swinburne argues that the only way to get "y=2x" out as the preferred explanation is by invoking the criteria of simplicity: "y=2x" is the simplest explanation. Of course, it is not as though the simplicity criterion is infallible: the next data point might not be captured by "y=2x". The simplicity of "y=2x" would have to be sacrificed for the sake of better accommodating the capturing-the-data criterion. But in our attempt to capture the new data, we would still look to the next simplest hypothesis; we will not have abandoned simplicity altogether. Although such a mathematical example is surely tidier than any example in philosophical theory evaluation, the simplicity criterion's role in resolving infinite indetermination in philosophical theory evaluation is still clear.

Such an argument—the simplicity criterion is necessary to make sense of our preferences among explanations, therefore it is a justified criterion—may seem simple, but Swinburne rests content there. Indeed, he argues against other attempts to justify the simplicity criterion. He argues that it cannot be understood merely as convenient, since we are concerned with truth not convenience when it comes to theory evaluation. He also argues that it cannot be justified a posteriori, by reference to what kinds of theories have been successful in the past. Successful theories in the past have not always been simple, nor can we pick out simplicity as a property of successful theories to be emulated without already invoking simplicity. He finally argues that the criterion cannot be justified analytically, arguing that any mathematical or logical attempt to do so will already be infected by simplicity. This leaves the simplicity criterion as a synthetic a priori proposition and Swinburne explicitly identifies it as such. Content to leave the justificatory question there, he says: "If simplicity could be justified further, it would derive that justification from some higher a priori criterion, and that one would be fundamental. We have to stop somewhere if particular judgments about the probability of theories and their predictions are ever to be justified, and the criterion of simplicity is, I suggest, the right stopping place."[68] He concludes his book by saying: "To summarise the claims in a nutshell: either science is irrational (in the way it judges theories and predictions probable) or the principle of simplicity is a fundamental synthetic a priori truth."[69] This strategy for justifying the simplicity criterion is similar to my preferred strategy, and I return to these issues in Chapter Five.

If simplicity can thus be shown to be a justified criterion of theory evaluation, then we will need an account of what makes one explanation or theory simpler than another. Swinburne thus goes on to give his preferred analysis of simplicity. His analysis is offered in explicit contrast to the analyses of Karl Popper and Elliot Sober.[70] Their analyses maintain that a theory is simpler if it has more content or higher falsifiability. But Swinburne insists, plausibly, that we can make comparative judgments of simplicity even among theories of equal

---

[68] Swinburne (1997), p. 51.
[69] Swinburne (1997), p. 56.
[70] See Popper (1959) and Sober (1975).

content (the mathematical example above is such a case). Though sometimes we do mean simplicity to be nothing more than content, there are enough important cases of simplicity that are not reducible to content to warrant a broader analysis.

Swinburne doesn't, however, offer a unified analysis of simplicity, but instead gives a list of "facets" of simplicity that are not reducible to content. These facets are as follows: (1) A theory is simpler if it posits fewer entities (this, Swinburne says, is Ockham's razor). (2) A theory is simpler if it posits fewer kinds of entities. (3) A theory is simpler if it contains fewer terms that require understanding of another term in order to grasp them (e.g. "grue"). This facet of simplicity can be rendered: "do not postulate underlying theoretical properties, unless you cannot get a theory which yields the data equally well without them." (4) A theory is simpler if it has fewer distinct laws. (5) A theory is simpler if its individual laws relate fewer variables. Finally, (6) a theory is simpler if it is mathematically simpler. Swinburne understands mathematical simplicity both as having fewer terms and as using more basic mathematical objects (lower numbers and easier operations).

Swinburne admits that having multiple facets of simplicity like this means that there is no unified analysis of simplicity. What's more, there must be trade-offs between the various facets of simplicity and there is no easy way to formalize how those trade-offs are to be made. But, Swinburne says, there is sufficient agreement on comparative judgments of simplicity, especially when holding some facets fixed, to make the account helpful. That is, though Swinburne's account of simplicity consists of a list, he maintains that complete, comparative judgments—holistic judgments about simplicity that weigh and balance all his facets of simplicity at once—are possible. Still, Swinburne's list of facets of simplicity is limited. It is particularly suited to scientific theories, as opposed to philosophical theories. Simplicity is a criterion of theory evaluation for both natural science and philosophy; a complete account of the criterion would speak to its application in both domains. What's more, having a mere list of separate facets of simplicity is itself un-simple; surely the simplicity criterion also means we should look for as simple or unified an analysis of simplicity as possible. And, because there is no unified analysis of simplicity, Swinburne can do little more than insist that holistic judgments are possible; he has no unified analysis to draw on in showing how complete judgments of simplicity are possible.

So, Swinburne offers an account of simplicity on which unified judgments of simplicity are possible: we can render one, all-facets-considered judgment when comparing two competing theories. Yet, his account only does this by reference to a non-unified analysis of simplicity, breaking down simplicity into a mere list of facets of simplicity. This tension will be a useful point of comparison when considering McAllister's subtly different account of simplicity. I motivate my own account of simplicity using that comparison. Thus, I now examine McAllister's account of simplicity.

Section 3: McAllister on Simplicity

McAllister's treatment of simplicity is doubly relevant to my project because he has a well-developed account of simplicity and because he argues for an aesthetic understanding of the simplicity criterion. His consideration of the criterion of simplicity comes within the context of examining how aesthetic criteria of theory evaluation could be a proper part of the rationalist enterprise of natural science; he numbers simplicity among those aesthetic criteria. Although his understanding of the aesthetic nature of simplicity will be different from mine, it is a helpful precursor to my argument.

To begin, I reconstruct McAllister's account of simplicity. McAllister's core claim about the simplicity criterion is the same as the core of Swinburne's analysis: there is no unity to simplicity. That is, there is not one type of simplicity, but rather a multitude of types of simplicity. There is not one simplicity criterion, but many simplicity criteria. McAllister provides a more technical analysis of this diversity than does Swinburne: First, McAllister argues, there is the division of form and degree of simplicity.

A form of simplicity is a property that a theory can possess that makes it apt to describe that theory as simple. Of these, McAllister gestures at several examples: 1) Explanatory simplicity: A theory can be aptly described as simpler if it explains a wider range of phenomena using the same explanation. 2) Ontological parsimony: A theory can be aptly described as simpler if it postulates a smaller number of entities. 3) Logical simplicity: A theory can be aptly described as simpler if it depends on a smaller number of independent postulates. 4) Numerical simplicity: A theory can be aptly described as simpler if it relies on (e.g.) simpler exponents or coefficients. Each of these, in turn, can be divided into several subforms. For instance, ontological parsimony can be understood as either qualitative or quantitative. Qualitative ontological parsimony requires a theory to postulate a smaller number of types of entities. Quantitative ontological parsimony requires a theory to postulate a smaller number of instances of the entities it posits. These "forms" of simplicity are obviously reminiscent of Swinburne's "facets" of simplicity.

The degree of simplicity is the degree to which a theory possesses the specified property of simplicity. This can be conceived as a scale: A simplicity criterion can require a theory to display the specified property to a high degree, or to a low degree, or anywhere in between. Naturally, we tend to think of simplicity criteria as requiring a high degree of simplicity. The simplicity criterion, even if it is properly specified into several distinct criteria that trade in different forms of simplicity, seems to require theories to be more simple, to show a higher degree of simplicity. But McAllister insists that this is not necessarily the case: we can conceive of criteria of simplicity that require a low degree of a familiar form of simplicity. For instance, one might prefer (in the spirit of Leibniz— whatever is possible is actual) theories to show ontological plentitude. That is,

one might deploy the criterion of simplicity specified to the form "ontological parsimony" and to the degree "low".

Alternatively, in contrast to McAllister's terminology, one might want to fix simplicity criteria at the degree "high" of whatever form of simplicity one prefers. One could then accommodate qualitative ontological plentitude by conceiving of a "baroqueness" criterion, which would be a perfect mirror to the simplicity criterion. The baroqueness criterion would trade in all the same forms as the simplicity, but would prefer theories to instantiate low degrees of each of these forms. Obviously, the simplicity and the baroqueness criteria would pull in different directions, and one couldn't maintain both of them as a justified criterion of theory evaluation for any given form of simplicity/baroqueness. But this is consistent with McAllister's picture, on which the various simplicity criteria—specified to a particular form and degree of simplicity—all pull in different directions. There is no guarantee that all forms of simplicity will prefer the same theory, especially when pulled to the opposite ends of the degree scale.

Though McAllister doesn't discuss it, it seems well within his spirit to alternatively think of simplicity criteria as fairly weak "checks" on theories. Such simplicity criteria would be tests that a candidate theory would have to pass, but passing such tests wouldn't be sufficient to establish the theory as preferable. These would be simplicity criteria that pick a particular form of simplicity and require theories to at least exhibit that form of simplicity to a moderate degree (with an ideal direction—high or low—also specified). Such simplicity criteria would be weaker than one that required the ideal high or low degree of simplicity of the relevant form, but could be deployed alongside other criteria of theory evaluation to come to a final determination. How much work we want a given simplicity criterion to do will affect how tightly we specify it.

However one prefers to slice these varieties of form and degree, McAllister's core claim remains: the simplicity criterion is underspecified among a great diversity of criteria. Each simplicity criteria, if it is to be properly specified, must include both a preferred form of simplicity and the degree of that form of simplicity it prefers. So, for example, one simplicity criterion could be: "prefer a theory that displays a high degree of the ontological parsimony form of simplicity." That is, prefer a theory that is highly ontologically parsimonious. This seems like Ockham's razor, and is close to what Swinburne identifies as Ockham's razor. McAllister is quick to point out, however, that Ockham's razor is already ambiguous between two interpretations. The classic formulation of the principle is, "Frustra fit per plura, quod potest fieri per pauciora" ("It is vain to do with more what can be done with less"). This can be read as the criteria: "prefer a theory that displays a high degree of the ontological parsimony form of simplicity." But it can also be easily read as the criteria: "prefer a theory that displays a high degree of the explanatory simplicity form of simplicity". This is McAllister's case in point: the simplicity criterion, even Ockham's razor, is underspecified. There are many simplicity criteria, each must be specified to a

particular form and degree of simplicity, and there is no guarantee that all these criteria will pick out the same theory in any given comparison.

So far, the analyses of Swinburne and McAllister seem very similar. But at this point they diverge. Unlike Swinburne, McAllister does not hold that unified, all-forms-considered judgments of simplicity are possible. That is, after dividing the simplicity criterion into the many criteria of the specified forms of simplicity, McAllister does not insist (as Swinburne does) that comparative judgments among the forms (or facets) of simplicity are possible. Indeed, McAllister's purpose is precisely to drive away from such holistic comparative judgments and to show that any comparison of simplicity simpliciter is dangerously underspecified and in need of disambiguation. The criteria of theory evaluation that are built around each form of simplicity all pull in different directions. As McAllister says: "The degree to which a theory displays one form of simplicity is uncorrelated with the degree to which it shows another."[71] This can seem like the more natural move to make, compared to Swinburne's insistence that comparative judgments across facets of simplicity are possible, given that there is no unifying analysis of the facets of simplicity. If there is nothing to unify the various facets or forms of simplicity, why should there be a way to make comparative judgments between them? Isn't the point of differentiating between various forms of simplicity precisely to discourage such overarching judgments?

But, perhaps surprisingly, McAllister does go on to offer a unifying account of all the various criteria of the forms of simplicity. McAllister's unifying claim is that all the criteria of simplicity are aesthetic. His account of simplicity falls within the larger context of attempting to show that aesthetic criteria can be part of a properly rationalist or empiricist picture of science. For that purpose, he develops an account of aesthetic criteria of theory evaluation, into which he plugs the simplicity criterion.

It is beyond the scope of my project to fully explore McAllister's account of aesthetic criteria of theory evaluation. This is because his account is expressly designed to fit the context of scientific theory evaluation as well as scientific revolution and it would take us too far afield into the philosophy of science to fully evaluate it.[72] But, roughly speaking, his account is as follows: McAllister holds that to pass an aesthetic judgment on an object (including an abstract object like a theory) is to evaluate it for beauty. (He grants that there are aesthetic properties other than beauty, but holds that beauty is the main aesthetic property we are concerned with in evaluating theories.) On McAllister's account, beauty is a value that is projected on to abstract objects like theories by (among others) scientists and scientific communities. Beauty is projected on to objects based on

---

[71] McAllister (1999), p. 113.

[72] I return to these notions—beauty, aesthetic properties, aesthetic judgment—and their relation to criteria of theory selection in the next chapter. There I am not constrained by analysis of McAllister's account, and explore these notions in greater depth as I develop my own account.

the aesthetic properties (like simplicity) that that object possesses. An aesthetic property is any property that can inspire the projection of beauty. Which properties will inspire someone to project beauty on to it will depend on the tastes and predilections of the projector: she will project the value of beauty on to those properties that she finds pleasing. McAllister holds that evaluating theories based on their beauty can be a proper part of rationalist or empiricist science through a process McAllister calls "inductive projection". Inductive projection works as follows: A scientist or a scientific community develops a taste for certain properties displayed by current successful theories. They project beauty on to those properties. They then go on to evaluate new or competing theories by reference to those properties. When a theory with very different or directly opposite aesthetic properties is clearly shown to be empirically successful, the scientific community undergoes a revolution and changes its aesthetic tastes to align with the aesthetic properties of the newly successful theory. The scientific community then projects beauty on to this new set of aesthetic properties and the cycle continues. This inductive projection of past aesthetic markers of successful theories on to future theories can be a proper part of a rationalist picture of science because it is an attempt (though an as yet unrealized attempt) to discover some correlation between a set of aesthetic properties and good scientific theories. McAllister remains agnostic about whether this pattern of inductive projection will ever be successful; it remains a possibility that no set of aesthetic properties will be correlated to good scientific theories. Though McAllister says very little about philosophical theory evaluation, presumably he thinks an analogous process is at work in philosophy.

Having given this account of aesthetic criteria of theory evaluation, as well as the account of simplicity above, McAllister moves quickly to connect the two. He says: "I think it is beyond doubt that, when scientists regard theories, some aesthetic pleasure is afforded to them by the theories' simplicity properties."[73] That is, it seems clear that simplicity is among those properties that inspire the projection of beauty. Of course, the taste for simplicity (or more precisely the taste for a particular form of simplicity) will shift over the course of inductive projection. The taste for one particular form of simplicity will give way to the taste for another particular form of simplicity as that second form is shown to be more empirically successful. It remains an open question which form of simplicity (if any) is correlated to the good scientific theories. It is clear from this, however, that there is, on McAllister's account, a unifying feature to all the forms of simplicity. They are all aesthetic properties and together comprise an important subclass of aesthetic properties.

Despite this unifying account, however, McAllister remains clear: he thinks that comparative judgments among the various forms of simplicity are impossible. This makes the account incomplete. Though McAllister's forms of simplicity are diverse, they are all still part of the same genus. That is, if each of the forms of simplicity is a particular sub-form of one, unified phenomenon

---

[73] McAllister (1999), p. 121.

(namely, simplicity itself), then some kind of comparison between the forms should be possible. This is especially pressing if we can give more content to how all these forms share a unified nature, as McAllister thinks we can. Such an explanation of how all the various forms of simplicity are united under one phenomenon should go some way toward explaining how comparative judgments between the forms are possible. Thus, though McAllister helpfully pushes us to disambiguate simplicity into its various forms, this does not amount to a complete account of the simplicity criterion itself.

McAllister's account, like Swinburne's, also is specifically suited to scientific theories, as opposed to philosophical theories. Again, simplicity is a criterion of theory evaluation for both natural science and philosophy; a complete account of the criterion would speak to its application in both domains. Swinburne and McAllister can hardly be blamed for this; they both write within the context of the philosophy of science. Nevertheless, simplicity is clearly also an important criterion of theory evaluation for philosophy. A complete account of the criterion will have the flexibility to apply to both domains.

Section 4: The Incompleteness of Simplicity

So, what conclusions have we reached? McAllister's account of simplicity does share one important feature with Swinburne's: it stresses the ambiguity of the simplicity criterion and the need to disambiguate it into various criteria built around the different forms of simplicity. But they each move from this "fact of diversity" to opposite conclusions: Swinburne concludes that there is no unifying analysis of the various facets of simplicity, but argues that comparative judgments across the facets of simplicity are possible. McAllister concludes that no comparative judgments across the forms of simplicity are possible, but argues that a unifying account of the various forms of simplicity is possible (namely, an aesthetic account of simplicity). But my main conclusion in this chapter is the point on which they agree: the simplicity criterion is underspecified. Both accounts, however, lack a clear means of adjudicating among the various forms of simplicity, and such adjudication is needed to fill out the simplicity criterion. Such adjudication among the forms of simplicity is needed if we are to have a complete account of the simplicity criterion and if we are to make sense of our use of the simplicity criteria in evaluating theories, scientific or otherwise.

As an example, we can return again to the case of who killed Mr. Boddy. Recall that Mr. Boddy was found dead in the library, that Colonel Mustard was found holding a candlestick with Mr. Boddy's blood on it, that Colonel Mustard stood to inherit Mr. Boddy's estate, and so on. It is clear that—given the facts to be explained and the limits on candidate explanations set by background considerations (statistics and human behavior)—the simplest explanation of Mr. Boddy's death is that Colonel Mustard killed Mr. Boddy. (Notice how when using

42

the simplicity criteria, we also use something like reflective equilibrium: in judging a theory to be simple or not we see how the candidate theory can fit, in a simple way, into a complete system of considerations. This makes it seem like simplicity is related to coherence: either a type of coherence or a part of how we make judgments of coherence or an alternative to coherence. This is striking, and I will return to this in later chapters.)

When we make the simplicity judgment about Colonel Mustard killing Mr. Boddy, we are considering several different forms of simplicity: We consider the simplicity of the number of entities invoked by the explanation: only Colonel Mustard is needed to explain Mr. Boddy's death, not additional malefactors who helped Colonel Mustard. We consider the simplicity of the number of types of entities invoked by the explanation: A blow to the head with the candlestick is all we need, rather than a blow to the head and a fatal disease. We consider the simplicity of the explanatory chain invoked by the explanation: Colonel Mustard did it to inherit the estate. An explanation that said that Professor Plum killed Mr. Boddy so that Colonel Mustard would inherit the estate so that Professor Plum could then steal the estate from Colonel Mustard would be less simple in this sense and we thus reject it. We are able to consider all these forms of simplicity at once and render an all-forms-of-simplicity-considered judgment about Colonel Mustard killing Mr. Boddy. No doubt there are more difficult cases, in which it is harder to make an all-forms-of-simplicity-considered judgment, but this does not undermine the importance of making such judgments, nor should we despair at the possibility of making such judgments.

Both Swinburne and McAllister's accounts fail to account for such all-forms-of-simplicity judgments. Swinburne wants a means of adjudicating between the various forms of simplicity, but doesn't give us an account of how such holistic judgments are possible. McAllister goes much of the way toward filling this lacuna—he provides a unifying account of simplicity. But he does not provide a means of making complete judgments of simplicity. (Indeed, it seems as though he doesn't want such judgments to be possible.) This means that both these accounts of simplicity have a lacuna remaining: Whether by choice or not, both accounts do not give a complete answer to how the simplicity criterion can help, all forms of simplicity considered, to adjudicate between competing theories.

This lacking is analogous to the lacuna in the notion of coherence in reflective equilibrium that I identified in the previous chapter: We lack a clear means of comparing the coherence of competing theories, and without such judgment the method of reflective equilibrium is incomplete. So, with these analogous lacunas in mind, I offer my own account of both reflective equilibrium and simplicity. My accounts fill both these lacunas; they provide an account of our judgments of coherence and simplicity that contributes to our understanding of those judgments. What's more, I fill both these lacunas with the same strategy. This is not merely fortuitous. Rather, it is a strength of the accounts:

they tie reflective equilibrium and simplicity together and deliver a lesson about the methods of theory evaluation and what makes theories good more generally.

As I indicated from the beginning, I accomplish all this by arguing that an aesthetic understanding of our judgments of coherence and simplicity can explain those judgments and thus complete the account of how these methods of theory evaluation work. This means that both reflective equilibrium and simplicity are partially aesthetic criteria of theory evaluation. Of course, such a strategy is in line with the account of simplicity that McAllister offered above. My account diverges from McAllister's—it relies on a different notion of aesthetic judgment. But it is inspired by McAllister's account. My account of simplicity combines the strengths of both Swinburne and McAllister's accounts: it unifies simplicity by reference to aesthetic judgment (a la McAllister), but still maintains that all-things-considered judgments of simplicity are possible (a la Swinburne).

Thus, in the next chapter I explore the notion of aesthetic judgment and its related concepts. Following that, I develop an account of both coherence and simplicity using that notion of aesthetic judgment. The aim is to fill the lacunas discovered in this and the previous chapter and so arrive at a more complete account of the method of reflective equilibrium and the criterion of simplicity.

CHAPTER FOUR

BEAUTY

Section 1: Understanding Methods of Theory Evaluation

In the previous two chapters I identified lacunas in the going accounts of reflective equilibrium and simplicity. Both of the methods were found to be incomplete because they both rely on notions—coherence and simplicity, respectively—that are unaccounted for. We have no account of coherence or simplicity that will help us understand those judgments such that we can make good judgments about which theories are more coherent or simple than others. In the case of coherence and reflective equilibrium, there is simply no account given to explain our judgments of coherence. In the case of simplicity, we did see accounts that explained the various forms of simplicity. But neither account we considered explained our judgments of simplicity simpliciter, all-forms-considered judgments of simplicity. Without understanding our judgments of coherence and simplicity, the methods that use them are incomplete, our understanding of them is incomplete. They do not, in the end, explain how we evaluate theories.

But one might retort: "Why do we need to understand our judgments of coherence or simplicity in order to complete the methods? Can't we just muddle through with our judgments of coherence and simplicity, and the methods that use them, even if we don't understand those judgments?" We can, of course, muddle through. But my point in articulating particular methods of theory evaluation is to clear up this muddle. That is, my purpose in examining these methods of theory evaluation is to improve our understanding of what makes theories good. As I explored in Chapter One, methods of theory evaluation can be understood as sorting better and worse theories or explaining what makes a theory better or worse. On either of these models, we will want to understand how using methods of theory evaluation leads us to better and worse theories. If we do not understand the judgment that a method of theory evaluation calls for, then we do not understand what makes a theory better. Theory evaluation will, of course, go on with or without any particular method of theory evaluation, or any particular account of those methods. But it would be a startling discovery to find that we do not understand the judgments that are at the root of reflective equilibrium and simplicity and thus that we do not understand how use of those methods leads us to better theories.

It seems that either we must conclude that the methods that employ the unexplained concepts of coherence and simplicity are incomplete methods of theory evaluation or we must find an account of coherence and simplicity. So

what might an account, an explanation, of our judgments of coherence and simplicity look like? The most obvious solution would be to articulate principles of coherence and simplicity.[74] But the difficulty of articulating such principles (or indeed any account whatever) is what got us into this problem in the first place. Is there any other way besides principles that we could explain our judgments of coherence and simplicity? Could judgments of coherence and simplicity be unprincipled and yet explicable?

Here at last we turn to beauty. I propose that judgments of beauty have precisely this structure—unprincipled, yet explicable, judgment—and that therefore judgments of coherence and simplicity should be understood as species of judgments of beauty. This fills out the needed explanation of our judgments of coherence and simplicity, without reference to any principles of coherence and simplicity. This fills in the lacunas in the accounts of reflective equilibrium and simplicity. The methods are thus completed: we have an explanation of how we should use the methods and how use of those methods leads us to better theories.

So, what is beauty and how can an account of it complete the methods in this way? Of course, the search for an account of beauty is as old as, say, the search for an account of what makes an action right. I do not fully answer the question "What is beauty?" A full answer to that question is beyond the scope of this project. Nevertheless, I lay out some important foundations for thinking about beauty and argue that these moderate foundations for a full theory of beauty are enough to do the work that I am interested in. Of course, even these moderate foundations for a full theory of beauty will not be utterly noncontroversial. But they are sufficiently noncontroversial so as to draw in engagement from many traditions of thinking about beauty (as well as common sense).

Beyond the moderate foundations for a theory of beauty that I sketch, I am agnostic about full theories of beauty. I do, however, lay out Mary Mothersill's full theory of beauty, to help fill out how understanding coherence and simplicity as species of beauty might look when plugged into a full theory. But it is worth emphasizing that I am not committed to any full theory of beauty, including Mothersill's, for the purposes of this project. The moderate foundations for a theory of beauty that I sketch are enough to show that coherence and simplicity should be understood as species of beauty and that this is sufficient to resolve

---

[74] What is and is not a "principle" is itself a difficult question and the subject of controversy. For now, I use "principle" to mean "a generalization that can yield a somewhat determinate judgment". I clarify this shortly. And even this minimal notion of "principle" might be objectionable: On certain conceptions, whether a thing is a "principle" or not might be distinct from how determinate it is. For instance, Kant's Categorical Imperative might be a paradigm principle. And yet, on certain readings of Kant, it might do *nothing* to determine whether particular actions are right or wrong.   The reader should keep these terminological difficulties in mind as I continue, somewhat uncomfortably, to use the term "principle".

the problems with the methods of reflective equilibrium and simplicity raised thus far.

Section 2: Beauty as Unprincipled yet Genuine Judgment

I begin with a few preliminary definitions and distinctions about beauty. These preliminary remarks are minimalistic. They home in on the concept of beauty, not provide a full theory of it.

Beauty is a value, akin to rightness or truth. A value is merely that which is good in some way. So we could think of beauty (perhaps like rightness and truth) as a species of good. That is, a thing's being beautiful is, all else being equal, good. This of course says nothing about how beauty is to be traded off against other goods or about what could swamp the goodness of beauty. Value judgments (at least as I use the term here) are claims about what is or is not of value. We judge things, for example, to be right, true, and beautiful. We make both comparative and non-comparative judgments about beauty (like other values).

I use the term "aesthetic judgment" to refer to judgments of beauty. The term "aesthetic" has a fraught history, of course, and I do not wade into those murky waters.[75] But, by definitional fiat, I use "aesthetic judgment" to mean "a judgment of the form 'Object O is beautiful'." One might also refer to this, in deference to the tradition of Kant and Mothersill that I draw on (as will become clear), as a "judgment of taste".[76] By this definition, thus, "aesthetic judgment" has nothing necessarily to do with artworks or with sensations, feelings, or pleasure (although, as we'll see, the connection to, say, pleasure might be quickly reestablished).

These sketches are meant to be non-controversial; they are simply given to help us home in on the concept of beauty. Such sketches form the minimal foundation for a full theory of beauty. I will now proceed to put forward some more substantive claims about beauty. Yet these claims do not yet amount to a full theory of beauty. I stay in this middle ground: My discussion of beauty is substantive enough to make the necessary connections to reflective equilibrium and simplicity, and yet not so substantive as to amount to a full theory of beauty. So, what more substantive claim about beauty could bring us back to the problems with the methods of reflective equilibrium and simplicity?

---

[75] For instance, there might be other aesthetic values besides beauty. "Artistic value" is a common candidate. But I am concerned only with beauty and will always mean "beauty-related" whenever I use the term "aesthetic". For more on the aesthetic as distinct from beauty, see Arregui and Arnau (1994), Beardsley (1973), Cavell (1965), Dickie (1974), Goldman (1990), Goodman (1978), McDowell (1983), Sibley (1959), Stolnitz (1960), Tolstoy (1899), Walsh (1974), and Wollheim (1968).
[76] See Kant (1790) and Mothersill (1984).

Consider the following two theses, drawn from Mary Mothersill and inherited from Kant:

First Thesis: There are no principles of taste.
Second Thesis: There are some genuine judgments of taste.[77]

The First Thesis states that there are no true, informative generalizations about what makes something beautiful. A principle of taste is an explanatory generalization about what is judged to be beautiful, a criterial feature of beautiful objects.[78] For example, "If an object has blue and red in it, then it is beautiful" and (more compelling) "If an object exhibits the golden ratio, then it is beautiful" are candidate principles of taste. The qualifier "interesting" in my gloss on the First Thesis above is important. There are, of course, circular principles of taste that are true. For example, "If an object is elegant, then it is beautiful," and "If an object has a pleasing arrangement of blue and red in it, then it is beautiful," are circular (Mothersill calls them "innocuous") principles of taste. They include beauty (or a cognate or derivative concept) in both the antecedent and consequent of the principle and so tell us nothing about how to make judgments of beauty. The First Thesis is not meant to deny that circular (innocuous) principles of taste are true; it only denies that any non-circular (interesting) principles of taste are true.[79]

The First Thesis can thus be framed as a dilemma for any candidate principle of taste: either it is interesting, but not plausibly true, or it is plausibly true, but innocuous (and so not actually a principle of taste). This indeed is Mothersill's main argumentative strategy for establishing the truth of the First Thesis: Take any candidate principle of taste that you like, scrub out any hint of beauty from the antecedent, and you will find (Mothersill predicts) that it is not plausibly true.[80] The principle of the golden ratio is a good example. It is perhaps the most plausible principle of taste on offer in the history of philosophy. And yet, once we are clear that the antecedent can contain nothing about beauty in it (the golden ratio must be a purely mathematical arrangement, not a "pleasing" mathematical arrangement), it falls flat. We can easily imagine many things that exhibit the golden ratio (simply an arrangement of mathematical ratios) that are not plausibly judged to be beautiful (Mothersill gives the example of her garage door).

---

[77] See Mothersill (1984), Chapters 4-6, and Kant (1790), 338-342. The labels "First Thesis" and "Second Thesis" are from Mothersill. For more on aesthetic principles and whether there are any, see Dickie (2005), Collingwood (1945), Goldman (2005), Mothersill (1989), and Zemach (1987).

[78] Notice that I am still glossing "principle" as a "generalization that can yield a somewhat determinate judgment". More on this in a moment.

[79] We could alternatively say that innocuous principles of taste are not *actually* principles of taste, and thus the First Thesis does properly deny that *any* principle of taste is true.

[80] See Mothersill (1984), pp. 126-131.

One difficulty in spelling out the exact nature of the First Thesis is worth examining: It is difficult to spell out exactly what counts as a "principle". Mothersill does not say much explicitly on this matter, but I draw out her implicit view and supplement it where necessary. I flagged the difficulty of using the term "principle" before. When I did so, I indicated that I had in mind the following: a principle is a "generalization that can yield a somewhat determinate judgment." Mothersill has something like this in mind as well. This is made clear by her consistent contrast between ethical principles and purported aesthetic principles.

Mothersill does not think that the analogue of the First Thesis is true for ethics. That is, she does think that there are ethical principles. But she does not flag this as a controversial view, which seems to indicate that she means this to be a harmless claim. She does not seem to mean, for instance, to claim that there are ethical principles that are algorithmic, or absolutely exceptionless (which would be controversial). This means that a generalization's being a principle is compatible with it have exceptions. This is perhaps surprising, since one might have thought that identifying exceptions to candidate aesthetic principles is how Mothersill argued against the existence of aesthetic principles (that is, argued for the First Thesis).

But this reveals that she has a more robust claim in mind when arguing for the First Thesis. She is not arguing that there are merely some extreme exceptions for any candidate aesthetic principle. Rather, she is arguing that any candidate aesthetic principle (unlike some plausible candidate ethical principles) will have rampant ordinary exceptions. For example, it is not as though the candidate principle "If an object exhibits the golden ratio, then it is beautiful" is merely false in certain extreme cases. Falsifying cases for the principle are as common as garage doors (again, to use Mothersill's example). The First Thesis is thus a stronger claim than it might first appear: one might have been prepared to accept it on the technicality that certain extreme exceptions undermine any candidate principle of taste that is meant to be exceptionless. But the First Thesis goes beyond this and asserts that any candidate principle of taste, even a moderate one that does not purport to rule out all extreme counterexamples, is easily shown to be false by ordinary counterexamples. Again, this is meant to be unlike the case in ethics, where it is plausible that at least some moderate principles (which are compatible with extreme exceptions) are true.

Thus, on Mothersill's usage, which I follow, a "principle" occupies a middle level of determinacy: it is not so determinate as to yield judgments exceptionlessly, but it is determinate enough such that ordinary and rampant exceptions are sufficient to undermine a generalization's being a principle. This is, admittedly, a hazy distinction: how many exceptions are sufficient to undermine a generalization's being a principle, are sufficient to make it not determinate enough? Nevertheless, Mothersill seems to think that the distinction is clear enough to make it apparent that there are no such principles for beauty, while it is plausible that there are such principles for ethics. For example, "It is right to keep one's promises" is a plausible moderate ethical principle: it is

immune to rampant ordinary counterexample (even if there are certain extreme counterexamples). In contrast, "It is beautiful to exhibit the golden ratio" is not plausible: it is susceptible to rampant ordinary counterexamples.[81]

We can connect this examination with my opening summation of coherence and simplicity and the problems that face the methods that use them. Recall that the problem with the methods of reflective equilibrium and simplicity is that they use forms of judgment—coherence and simplicity—that lack an account, that are unexplained. The main we reason we felt this lack of explanation is because we found no principles—using the sense of "principles" I just laid out in analyzing Mothersill's First Thesis—of coherence and simplicity. So, again, the problem is: If we have no principles—generalizations that can yield judgments free of rampant ordinary counterexamples—of coherence and simplicity, how can we account for those judgments in a way that would complete our understanding of the methods of reflective equilibrium and simplicity? Can we make unprincipled judgments about better and worse theories?[82] If so, how do we explain those unprincipled judgments? Let's turn now to Mothersill's Second Thesis.

The Second Thesis states that some judgments of beauty are genuine, that is, true. That is, the Second Thesis claims that some judgments of the form "Object O is beautiful" are true. This seems like a more harmless claim than the First Thesis and indeed it is. Any discussion that takes beauty seriously presupposes that some of our judgments about it are on the right track. Importantly, the Second Thesis does not affirm the truth of any particular judgment of taste (which would be more controversial and, in any case, not

---

[81] We could put it another way: "Principles" in Mothersill's sense include a *ceteris paribus* clause. Extreme counterexamples are *not* sufficient to undermine such a principle because they will be cases where "all else is *not* equal". But if there are counterexamples *even when* "all else *is* equal"—ordinary counterexamples—then that will be enough to undermine such a principle. The First Thesis claims that there are *no* such *ceteris paribus* principles for beauty (where in contrast there *are* such *ceteris paribus* principles for ethics). Mothersill consistently maintains this contrast case with ethics, maintaining that there are *ceteris paribus* principles in ethics. While this is a widely shared view, it is worth noting that it is not *entirely* uncontroversial. For example, Jonathan Dancy maintains that there are not even such moderate ethical principles, that ethical judgments are not principle-governed in any way. When we remember that principles, at least as I have defined them here, do not have to be *exceptionless*, but merely generalizations that hold over normal circumstances, it is clear that this is the minority position in ethics.

[82] I continue to use, somewhat uncomfortably, the term "principle" throughout. I consistently mean this moderate understanding of "principle": a generalization that is free from rampant ordinary counterexamples (though not necessarily extreme counterexamples). I mostly speak of judgments (like judgments of beauty, coherence, and simplicity) that are "unprincipled". This means judgments that *lack* even such moderate generalizations. That is, an unprincipled judgment is made without reference to any generalizations that are free from rampant ordinary counterexamples.

pertinent to general theorizing about beauty). It only affirms that some unspecified judgment of taste is true. Pick whatever judgment of taste you find to be most plausible ("If anything is beautiful, then this is surely beautiful…") and that is evidence for the Second Thesis. The Second Thesis asserts that no one is willing to give up all her judgments of taste.

Though I followed Mothersill's terminology and way of framing the two theses, they come from Kant's Critique of Judgment (as Mothersill acknowledges). Kant's way of putting the First Thesis is that "there is no disputing about taste" or that "judgments of taste are not determined by a concept".[83] Indeed, he maintains that these features of judgments of beauty are necessary and that they can be known a priori. (Mothersill only argues for the First Thesis by showing that the prospects for any candidate principle of taste are dismal; she does not argue that it is a necessary truth.)

Mothersill suggests that it was Kant's greatest insight in theorizing about beauty to recognize that a theory of beauty must capture and explain both theses.[84] How are some judgments of taste genuine if there are no principles of taste with which to adjudicate those judgments? This, Mothersill claims following Kant, is the chief puzzle of theorizing about beauty. It is at this point that a full theory of beauty would step in and explain how the two theses cohere, how judgments of taste are possible (to use the Kantian term) given that there are no principles of taste (again, using "principles" in the sense laid out above).[85]

But, as I've said, my purpose is not to give a full theory of beauty. My purpose is to use the concept of beauty to fill in the explanatory lacunas remaining in the methods of reflective equilibrium and simplicity. We now have enough of a grip on the concept of beauty to do this: A distinctive feature of aesthetic judgment is that it is not governed by principle, and yet that it is nevertheless possible. A full theory of beauty is needed to fully explain how this is possible; but the foundation of a theory of beauty that I laid out above is enough to establish that it has this distinctive character. And that it has this distinctive character is enough to bring us back to simplicity and reflective equilibrium. In examining those methods we concluded that simplicity and coherence (part of the method of reflective equilibrium) also have this character: judgments about them are surely possible, and yet such judgments are not governed by principles (or any account whatever).

---

[83] See Kant (1790), 338.

[84] For more on Kant's theory of beauty, see Allison (2001), Ginsborg (2013), Guyer (1979) and (1993), and Kant (1790).

[85] See Mothersill (1984), p. 169.

Section 3: A Match Between Beauty and Coherence/Simplicity

So, some moderate claims about beauty get us to the conclusion that a distinctive character of judgments of beauty is that they are unprincipled and yet can be genuine. Judgments of coherence and simplicity, according to my analysis above, also have this character: we have no principled account of either coherence or simplicity, and yet the methods of reflective equilibrium and simplicity require that we be able to make genuine judgments of coherence and simplicity in order for the methods to be complete. There is thus an analogy between the structure of judgments of beauty, on the one hand, and judgments of coherence and simplicity, on the other hand. It is worth recalling some of the conclusions from the chapters on reflective equilibrium and simplicity to further establish this analogy.

In the case of reflective equilibrium, there was simply no account on offer for coherence, a critical piece of the method. Rawls and his commentators offer no principles of coherence, nor are such principles even flagged as forthcoming or possible. Coherence is only glossed as "fitness" between the various beliefs that compose a reflective equilibrium. Yet, coherence is a critical piece of the method: it provides the normative pressure to revise beliefs in light of others in the reflective equilibrium. So, if the method of reflective equilibrium is to be complete—able to terminate in judgments about which theories are better than others—it must be the case that judgments of coherence are possible. So, judgments of coherence are possible (or at least this is presumed by the method of reflective equilibrium) and yet they are not governed by principles (or least we have an utter lack of promising principles of coherence).

In the case of simplicity, there was more of an attempt to give principles of simplicity. Both Swinburne and McAllister distinguished various types of simplicity and tried to articulate a principle for each of these types of simplicity: e.g., "If a theory postulates fewer entities, then it is more ontologically parsimonious." Since the antecedent of this principle does not contain a reference to simplicity, the principle is non-innocuous. Such principles could be given for each of McAllister and Swinburne's types of simplicity. But these principles are not principles of simplicity simpliciter: they do not purport to articulate a principle that can be used to judge when something is all-things-considered simpler. Indeed, Swinburne and McAllister are driven to differentiate between the various forms of simplicity precisely because they despair at the prospect of a principle of simplicity itself. The various forms of simplicity all pull in different directions, so there is not any unifying principle for simplicity itself. Swinburne concluded that despite this lack of principles, judgments of simplicity itself (that is, judgments that cut across the various forms of simplicity and declare something to be all-things-considered simpler) are nevertheless possible. He concludes this despite the fact that he concedes that no unifying principle of simplicity is possible. McAllister goes the other way: he concludes that because

there are no unifying principles of simplicity, all-things-considered judgments of simplicity are not possible.

I argued in the previous chapter that Swinburne is right to maintain that holistic judgments of simplicity—judgments of simplicity simpliciter that cut across the various forms of simplicity—are possible. We clearly make such holistic judgments, not only about theories or explanations, but about people, systems, strategies, artworks, and natural objects. (This fact alone is enough to show that Swinburne and McAllister's analyses of simplicity are incomplete: they only apply to theories, much less other objects that we judge to be more or less simple.) But McAllister is right to demand more explanation if we are to maintain that holistic judgments of simplicity are possible without a unified principle of simplicity. How are such holistic judgments of simplicity possible if there are no unifying principles of simplicity? As should be clear, this is precisely the same question that faces aesthetic judgment.

I described judgments of simplicity simpliciter as "holistic" judgments. I refer to them this way because such judgments are supposed to take into account all the various forms of simplicity and deliver an all-things-considered judgment of simplicity. The distinctive character of judgments of beauty that I am focusing on can be framed this way as well. When making judgments of beauty we recognize that there are many factors to be considered. Suppose you are judging a painting on whether or not it is beautiful. You consider the way color is used, you consider the way the subject is arranged, you consider the texture of the materials used, you might even consider the painting's frame and contextual features like lighting. All of these things could pull in different directions as to whether or not the painting is, all-things-considered, beautiful. And yet (according to Mothersill's Second Thesis), we can (at least sometimes) make genuine all-things-considered judgments of beauty. We could articulate various "forms" of beauty and attach a principle to each of them. But we still would lack a principle of beauty simpliciter, a principle that could guide us to all-things-considered judgments of beauty. And, given that we make genuine all-things-considered judgments of beauty, the puzzle would remain. The same holds true of coherence: We could articulate various "forms" of coherence. But we make all-things-considered judgments of coherence that cut across these various forms. The distinctive character of judgments that I am focusing on is unprincipled, but genuine judgment. Such judgments are the same as all-things considered judgments that are made without reference to a principle that is capable of adjudicating among the various forms of the relevant phenomenon.

So, there are genuine judgments of beauty, coherence, and simplicity. Such judgments can be comparative or non-comparative and can render a holistic judgment over considerations that pull in different directions. These judgments can do this without ever invoking a principle to determine the judgment. The question still remains how these judgments are possible. I did not (and will not) answer that question. I am only arguing that judgments of beauty, coherence, and simplicity share a distinctive characteristic.

Given this analogy between the structure of judgments of beauty, coherence, and simplicity, I suggest that coherence and simplicity are species of beauty. That is, judgments of coherence and simplicity are types of aesthetic judgment (in the same way that judgments of, say, elegance or gracefulness are types of aesthetic judgment). To put the argument flat-footedly: "Judgments of beauty, coherence, and simplicity all share the same distinctive feature. Beauty is the broadest of the three categories.[86] So, beauty is a genus of which coherence and simplicity are species." This argument is not, of course, deductive. I did not show that it is necessary that coherence and simplicity are species of beauty. The argument is, rather, that the best explanation of the analogy between beauty and coherence/simplicity is that the two are species of beauty. Given that there are clearly judgments of beauty that are not about coherence or simplicity, the relation between beauty and coherence/simplicity must be a genus/species relation. That is, the analogy suggests that all judgments of coherence or simplicity are judgments of beauty, but it is clear that not all judgments of beauty are judgments of coherence or simplicity. (Of course, what makes this genus/species relation seem like the best explanation of the analogy is that it is the simplest or most fitting explanation. This is not a vicious circle, since I did not suggest that we suspend use of the simplicity criterion or the method of reflective equilibrium, only that they require a more complete explanation.)

But one might wonder: Aren't there other kinds of judgments that are unprincipled, and yet possible? Why think that coherence and simplicity are species of beauty instead of some other type of judgment that shares that same characteristic? First, this objection does not have much force until a specific candidate type of judgment, which has this same characteristic, is identified. I find it difficult to identify any such type of judgment besides judgments of beauty. Remember, the chief contrast to judgments of beauty, which are unprincipled, is ethical judgments, which are principled.[87] When we are clear on what it means for a judgment to be unprincipled, I think it is difficult to find examples of unprincipled judgments, other than judgments of beauty. Second, it will be even more difficult to identify another candidate type of judgment than it might seem. This is because the candidate type of judgment must not only share the

---

[86] To say that beauty is the broadest of the three concepts is merely to say that it has much more diverse application: a great many kinds of things are beautiful and for many different kinds of reasons. It is clear that coherence and simplicity *could* be a species of beauty, but it would be nonsensical to say that beauty could be a species of coherence or simplicity: beauty is simply too big of a concept to be subsumed under coherence or simplicity.

[87] This way of framing things is still assuming that particularism about morality is false: that there are some generalizations about morality that hold over normal circumstances (derived, perhaps, from a master principle of morality like the Categorical Imperative or the Golden Rule). This is, as I noted above, not entirely uncontroversial, but is widely believed.

distinctive feature—unprincipled, and yet possible—but it must also be big enough to be a genus into which coherence and simplicity can fit. That is, the conclusion that coherence and simplicity are species of beauty is not only vindicated because they all share the distinctive feature. It is also vindicated because beauty is a broad enough category of judgment to be a genus into which coherence and simplicity can fit. Beauty is a broad genus of judgment and we are already familiar with its having several species: elegant, graceful, striking, harmonious, bold, pretty, charming, etc. Coherence and simplicity are of the right conceptual scope to fit into this list of species of beauty. Thus, a candidate alternative type of judgment must both be unprincipled and possible and be a broad enough genus of judgment such that it can encompass coherence and simplicity alongside its other species. I am skeptical that such a type of judgment—besides beauty—can be identified.

Nevertheless, one might think that the analogy between the structure of judgments of beauty, coherence, and simplicity is still not enough to establish that coherence and simplicity are species of beauty. If this is so, then simply consider how often they are treated as species of beauty. There is considerable discussion in the philosophy of science on how aesthetic judgment relates to scientific theory evaluation.[88] (Because my project here is aimed at philosophical theorizing in particular, and theorizing simpliciter in general, it is different and broader than these discussions in the philosophy of science.) In these discussions, coherence and simplicity are frequently explicitly counted among aesthetic considerations that are used to evaluate scientific theories.[89] Or consider our aesthetic judgment of artworks. We frequently use coherence as part of aesthetic judgment of artworks. For example, "All the pieces of the plot come together really well, so the play is quite beautiful," or "The symphony succeeds because the theme comes full circle by the end," or "Because the dark colors fit with the dramatic nature of the scene, the painting hangs together." (It's clear enough that these statements use the notion of coherence, or fitness, to make a judgment of beauty. I am not claiming that we slavishly use the word "coherence" in a large proportion of our aesthetic judgments.) We also use simplicity in our aesthetic judgment of artworks. For example, "The clean colors are what makes the painting so striking," or "The piece is beautiful because of its uncomplicated melody," or "The purity of the author's voice makes the poem excellent." (Again, it need not be the case that the word "simplicity" appears constantly.)

I am, of course, not suggesting that coherence and simplicity exhaust the space of what is beautiful. Coherence and simplicity are two among many

---

[88] For discussions of the role of aesthetic judgment in science, see Chandrasekhar (1987), Curtin (1982), Engler (1990), Kivy (1991), Martin (1989), McAllister (1989) and (1999), Weyl (1952), and Zemach (1986).

[89] See, for example, Chandrasekhar (1987) and Engler (1990). They explicitly identify coherence and simplicity as species of beauty that are relevant to scientific theory selection.

species of beauty (elegant, striking, harmonious, etc.). Nor am I suggesting that coherence and simplicity can be used as principles in judgments of beauty. Remember: any candidate principle of taste will be either true but innocuous or interesting but false. I am arguing that coherence and simplicity are species of beauty. If that is so, then the notion of beauty is already in the antecedent of the principle, "If a thing is coherent or simple, then it is beautiful." Such a principle is circular and innocuous (if coherence and simplicity are species of beauty). If simplicity and coherence are not species of beauty and are scrubbed of any aesthetic content, then the principle is obviously false. (Of course, I am arguing that scrubbing coherence and simplicity of any beauty-related content is impossible, given that they share a fundamental distinctive feature of beauty.)

Identifying coherence and simplicity as species of beauty also explains the connection between coherence and simplicity. We saw hints of this connection when we looked at the role that judgments of coherence and simplicity play in constructing and selecting theories. Earlier I noted that the two methods each are primarily connected with one of these two stages of theory evaluation: the method of reflective equilibrium (and thus judgments of coherence) is particularly suited to building theories and the criterion of simplicity (and thus judgments of simplicity) is particularly suited to selecting theories. But, I also noted that this tendency is not absolute: both methods are used at both stages. For example, simplicity is relevant when trying to use reflective equilibrium to build theories because part of what is for a theory to be coherent is for it to be simple. If a theory "fit" with all the relevant considered judgments, but was hopelessly complicated, we would not judge this theory to be coherent. Thus, simplicity and coherence seem closely related somehow; they can sometimes seem like two parts of the same judgment. We can now explain this relationship between simplicity and coherence: judgments of simplicity are often relevant to judgments of coherence (and vice versa) because they are both part of the same family of judgment: beauty. Because they are both species of beauty, they will often be relevant to each other, in just the same way that, say, elegance and harmoniousness are relevant to each other. The fact that identifying coherence and simplicity as species of beauty has this explanatory benefit is, in turn, a reason to be believe that coherence and simplicity are species of beauty.

So, to reiterate, judgments of beauty, coherence, and simplicity share an analogous structure. They are each unprincipled and yet can deliver genuine judgments: There are true judgments of beauty, coherence, and simplicity. Such judgments can be absolute or comparative and can render a holistic judgment over considerations that pull in different directions. These judgments can do this without ever invoking a principle to determine the judgment. Beauty is a broad genus, with many different species, so simplicity and coherence are of the right conceptual scope to be counted among beauty's species. What's more, coherence and simplicity are frequently cited in ordinary judgments of beauty about theories and artworks. And, what's still more, this genus/species relation explains the connection between simplicity and coherence and why they

56

sometimes seem like two parts of the same judgment. This is enough to establish that coherence and simplicity are species of beauty: judgments of coherence and simplicity are types of judgments of beauty.

Section 4: Completing Our Understanding of the Methods

I concluded that coherence and simplicity are species of beauty. What does this tell us? How does this solve the problems I raised in my analyses of the methods of reflective equilibrium and simplicity? Recall that the problem I identified with each of the methods was that they rely on an unaccounted for notion to complete the methods. In order for the method of reflective equilibrium to get us to a preferred theory (or a reasonably small set of preferred theories), the method instructs us to make judgments of coherence, to judge how coherent one theory is as compared to another. Likewise, in order for the simplicity criterion to get us to a preferred theory, the criterion instructs us to make judgments of simplicity, to judge how simple one theory is as compared to another. The nature of these judgments of coherence and simplicity remains unexplained in the going analyses of these methods, and so the analyses of these methods remain explanatorily incomplete: we do not have a complete explanation of how these methods terminate in better theories.

Now, I argued that judgments of coherence and simplicity are types of judgments of beauty. This tells us something about these judgments and so tells us something about the methods that require them. It provides more of an explanation of judgments of coherence and simplicity by locating them in a more familiar category. Of course, without a full theory of beauty, we still do not have a full account of how these judgments work. But the lack of such a full theory of beauty does not undermine our confidence in our judgments of beauty, and judgments of coherence and simplicity can share in that presumptive position by being identified as species of beauty. Judgments of coherence and simplicity are no longer utterly mysterious and unaccounted for, but are categorized in such a way that we can have a rough grip on them and we can know how to proceed in thinking about them.

And if judgments of coherence and simplicity inherit a presumptive status in virtue of being identified as types of judgments of beauty, then this serves to explanatorily complete the methods of reflective equilibrium and simplicity; it completes our understanding of the methods. Rather than relying on an unexplained type of judgment, the methods instruct us to make aesthetic judgments. That is, the methods instruct: "One can arrive at a preferred theory (or set of theories) by making judgments of coherence or simplicity, which are

types of judgments of beauty."[90] That final clarification, the explanation of coherence and simplicity that I am arguing for, serves to provide one deeper level of explanation on how the methods work. It puts all the methods' cards on the table and makes it clearer exactly what the methods are calling for in instructing us to make judgments of coherence and simplicity. This is enough to complete the methods, to explain how they are meant to get us to terminate on better theories. This is not to say that this explanation is a silver bullet for solving any problem the methods might face. One might still object to either method on entirely independent grounds. Indeed, it opens up new frontiers of problems for the methods since there remain large and difficult questions about the nature of judgments of beauty. But it does explain what kind of judgment the methods are calling for in evaluating theories. It thus completes the methods in an important sense: it gives a complete account of what judgments the methods instruct us to make in order to terminate in a better theory (or set of theories).

 To see this, consider again the case of Mr. Boddy, found dead in the library with Colonel Mustard holding the bloody candlestick. We saw earlier how we can use the methods of reflective equilibrium and simplicity to explain how and why we should settle on the theory that Colonel Mustard killed Mr. Boddy: That theory coheres (fits) with all the information we have and the relevant background considerations; that theory is also the simplest available theory (where the background considerations determine which theories are "available"). That is, we saw that we used judgments of coherence and simplicity to settle on the theory that Colonel Mustard killed Mr. Boddy. Now consider how we would use judgments of beauty if we were asked: Which theory of how Mr. Boddy died is the most beautiful or the most elegant? Though we could cite no principle of beauty, it is clear that we would still settle on the theory that Colonel Mustard killed Mr. Boddy as the most elegant theory, and that we would cite the same considerations: fitness with the available information and simplicity.

 Or consider going the other way: Suppose someone asks you which of two paintings is more beautiful. They are both abstract paintings composed solely of arrangements of color. When you make your judgment of beauty, without reference to any principle, you will say things like: "This one is more beautiful because the colors come together to create a distinctive mood," or "This one is less beautiful because the colors clash," or "This one is more beautiful because the arrangement flows nicely," or "This one is less beautiful because the arrangement is haphazard". Now suppose someone asks you which of the same two paintings is more coherent or simple. When you make your judgment of coherence or simplicity, you will say similar things. This will be the same kind of judgment you make when judging the theory that Colonel Mustard killed Mr. Boddy to be the best theory: it considers the various features of the object,

---

[90] Of course, in the case of reflective equilibrium, the judgment of coherence is only *part* of the method, so the method's instruction to make a judgment of beauty will likewise be only *part* of the method.

makes no reference to any principle, yet renders an all-things-considered judgment about the goodness of the object.

Comparing our judgments of coherence/simplicity and our judgments of beauty in these ways highlights how similarly structured they are—such judgments make no reference to principles, yet they are clearly possible. They consider the features of the judged object and render an all-things-considered judgment, where none of those features are criterial features for the judgment. These cases also show that it is plausible that coherence/simplicity are species of beauty: we often cite our judgments of coherence and simplicity when we are asked to make judgments of beauty. These cases—both the theory of who killed Mr. Boddy and the abstract paintings—also indicate that coherence and simplicity are themselves closely related, as we've seen hints of throughout my analysis.

Thus, my aesthetic understanding of the methods of reflective equilibrium and simplicity—make judgments of coherence and simplicity, which are types of judgments of beauty—completes the methods because it locates our judgments of coherence and simplicity alongside other judgments with which we are familiar: judgments of beauty. This explanatory completion of the methods does not necessarily immediately improve our ability to make judgments of coherence and simplicity. It obviously does not improve our ability to make judgments of coherence and simplicity by providing principles of coherence and simplicity (which would be the most obvious way to improve those judgments). But it does improve our understanding of the methods of reflective equilibrium and simplicity by helping us see—by reflecting on our other judgments of beauty—what kind of judgments we should be making. When we make judgments of coherence or simplicity when using reflective equilibrium or simplicity to evaluate philosophical theories, we can reflect on how we would make such judgments about, say, artworks. Such comparisons, used to help us focus on the type of judgment we are making and to help us use that power of judgment as best we can, will improve our understanding of the methods of reflective equilibrium and amounts to a completion of the methods. This may, over time, improve our judgments of coherence and simplicity.

To put it another way, we have found a way to explain our judgments of coherence and simplicity without any reference to principles. This increases our understanding of the methods that require those judgments. Again, this strategy for explaining our judgments of coherence and simplicity does not rely on principles (even moderate principles) of coherence and simplicity (which is the most obvious way to explain a judgment). Rather, it locates those judgments in precisely that category of judgment that enjoys genuine judgment without principles: judgments of beauty. Judgments of beauty, by Mothersill's two theses, have the special character of being possible (able to arrive at genuine judgments) without reference to any principles. If judgments of coherence and simplicity are types of judgments of beauty, then they can occupy precisely that

middle space—unprincipled, yet possible—they need to if we are to have a complete understanding of the methods of reflective equilibrium and simplicity.

To put it yet another way, the solution I am arguing for is to subsume the puzzle about judgments of coherence and simplicity—"What exactly are judgments of coherence and simplicity?"—under the larger puzzle about judgments of beauty. This, in some sense, only pushes the puzzle back one step. We still want a full theory of beauty to explain precisely how judgments of beauty are possible: how genuine, but unprincipled judgment is possible. But this is a fruitful step to take in thinking about coherence and simplicity. We have learned that the puzzle of coherence and simplicity is not an isolated puzzle, and we can gain more confidence about our judgments of coherence and simplicity by learning that they are in the same boat with other judgments that we are confident about: judgments of beauty.

To further clarify my claim that we can complete the methods of reflective equilibrium and simplicity by understanding coherence and simplicity as species of beauty, it will be fruitful to briefly consider what this claim denies. It denies that an explanation of coherence and simplicity that does not invoke beauty will ever be sufficient to explanatorily complete the of methods. That is, it denies that we will find principles of coherence or simplicity that will explain our judgments of coherence and simplicity. That is, it denies that it is fruitful to completing our understanding of the methods of reflective equilibrium and simplicity to try to develop freestanding accounts of coherence and simplicity (as perhaps Rawls, Daniels, Swinburne, and McAllister have tried, respectively, to do). So, if we should not look to principles or to developing a freestanding account of coherence and simplicity in trying to complete the methods, how should we proceed in applying the methods? How can we use aesthetic judgment—either of the coherence type or the simplicity type—to work out the methods of reflective equilibrium and simplicity in order to get to better philosophical theories?

Consider how we use aesthetic judgment in other domains and how we sharpen our aesthetic evaluations of aesthetically better and worse objects. In the case of using aesthetic judgment to evaluate artworks or the natural world, there are many resources for sharpening our aesthetic judgment. One can be trained, for example, to listen attentively to music, to understand a complicated plot structure, to appreciate the mechanics of an ecosystem. By having experts point out easily missed details to us, by repeated exposure to an object, and by attentiveness we can improve our aesthetic judgments. With such training we are able to make better and more precise judgments about what is more and less beautiful.[91] My strategy for completing the methods of reflective equilibrium

---

[91] I am consciously echoing Sibley's famous and excellent account of how we can improve our aesthetic judgment. See Sibley (1958). Although I have not framed my discussion using his terminology, his account of "aesthetic concepts" is clearly a close cousin to Kant/Mothersill's antimony of taste: judgments of beauty are unprincipled, yet possible. In his terminology: Non-aesthetic properties (properties that are perceivable

proposes that we should develop the same rigorous practice of aesthetic judgment when it comes to philosophical theories. That is, if my strategy for completing the methods is correct, and if we wish to use reflective equilibrium or simplicity, we must train ourselves to aesthetically evaluate philosophical theories.

Now, generally improving aesthetic judgment in one domain helps in another domain. For example, improving one's evaluation of literature helps, in a modest way, to improve one's evaluation of paintings.[92] This means that it will be modestly beneficial to philosophical theory evaluation to work at improving aesthetic evaluations of, say, artworks. But, on the other hand, the details of aesthetic judgment are also specialized to particular domains. For example, there are significantly different kinds of details to watch for in evaluating a play as compared to a painting. This means that we will have to develop canons (but not principles) and techniques of aesthetic judgment that are particularly suited to philosophical theories. What are the details to watch out for in looking for a beautiful philosophical theory? An incomplete list of possible such details would be: How does a theory fit with other plausible theories? What kind and how many principles does a theory use? What kind of symmetries does a theory exhibit? What kind and how many disjunctions does a theory use? What kind and how many modal properties does a theory use? All of these seem relevant to an aesthetic evaluation of philosophical theories, but it is far beyond my scope here to fill all these out (much less the many others we could list). Developing such a canon of aesthetic judgment for philosophical theories would be a difficult and somewhat strange task.[93] But it is what my conclusion—judgments of coherence and simplicity should be understood as aesthetic judgments in order to complete the methods of reflective equilibrium and simplicity—recommends.

Of course, if all this is on the right track, then this means not only that we can sensibly label judgments of coherence and simplicity as "aesthetic" (remember, I am using "aesthetic" only to mean "beauty-related"). We can label them as such because they are types of aesthetic judgment. But it also means that we can sensibly label the methods of reflective equilibrium and simplicity as "aesthetic". We can label them this way because they each instruct us to use a type of aesthetic judgment to arrive at better theories. In the case of the method of reflective equilibrium, the label "partially aesthetic" would be more appropriate, since the judgment of coherence (which I identified as aesthetic) is only one part of how the method works.

---

through normal sense perception) are not *conditions* for aesthetic properties, yet we nevertheless have the power to perceive aesthetic properties.

[92] Fully vindicating this phenomenon and explaining how it works is a difficult puzzle in aesthetics.

[93] "What exactly is a 'canon'?" is itself a difficult question. It is part of the question of how we should go about systematizing our aesthetic judgments without recourse to principles. This is itself part of the daunting task of providing a full theory of beauty.

To discover that two very popular and powerful methods of theory evaluation are aesthetic (or partially aesthetic) is a surprising conclusion. Philosophers do not usually think of judgments of beauty as being relevant to philosophical theory evaluation. And yet, if my arguments here are correct and if the methods of reflective equilibrium and simplicity are good and justified methods of theory evaluation (which it seems to me that they are, though I did not directly argue for this), then we do sometimes use aesthetic judgment to select philosophical theories. Of course, aesthetic methods of theory evaluation are not unfamiliar to philosophers of science in thinking about scientific theory evaluation (though the existence and justification of such aesthetic methods of scientific theory evaluation is controversial).[94] But we are here led to aesthetic methods of philosophical theory evaluation, a beast of a different order. Of course, there will be interesting and difficult questions as to how aesthetic methods of scientific theory evaluation relate to aesthetic methods of philosophical theory evaluation (I think they are contiguous, though I offer no argument for this). And if my arguments are successful, then this provides support for philosophers of science who argue that there are justified aesthetic methods of scientific theory evaluation.

We can again further clarify what I am claiming—the methods of reflective equilibrium and simplicity should be understood as (partially) aesthetic—by considering what the claim denies. It denies that an account of these methods that makes no reference to beauty will be explanatorily complete. That is, an account that explains the judgments about better and worse theories that result from using the methods must make reference to judgments of beauty. This thus denies that aesthetic judgment has no place in any of our methods of theory evaluation. Of course, I am not claiming that all methods of philosophical theory evaluation rely solely on aesthetic judgment, that aesthetic judgment is all there is to evaluating philosophical theories. I only concluded that aesthetic judgment is one important tool we use to select philosophical theories and thus I deny any picture of philosophical theory evaluation on which aesthetic judgment is completely excluded.

Although I have analyzed both reflective equilibrium and simplicity in the general terms of methods of theory evaluation, recall that I also distinguished between two stages of theory evaluation: theory construction—for which reflective equilibrium is a particularly important method—and theory selection—for which simplicity is a particularly important criterion. One might have thought that judgments of beauty could be used to select between theories, but only as a tiebreaker between two otherwise equally good theories. I am now in a position to respond to this. First, it will depend on how strictly we understand "tiebreakers": criteria that can be used to select between otherwise equally good theories. If one thinks that the tiebreaker of beauty can be used only when there

---

[94] See McAllister (1996) for an overview of the debate on aesthetic methods of scientific theory selection as well as his own candidate theory of aesthetic methods of scientific theory selection.

is not even an iota of difference in quality between competing theories, then this is far too restrictive of judgments of beauty. This is because we use the criterion of simplicity (and thus judgments of beauty) to select between theories even when they are not exactly otherwise equally good. If a theory is marginally worse than its competitors on all other metrics, but is far simpler, we would select that theory as the best. If a criterion's being a tiebreaker means that it can only come into play after all other metrics have been completely exhausted, then the criterion of simplicity (and thus judgments of beauty) is not merely a tiebreaker. We could speak of the criterion of simplicity (and thus judgments of beauty) as a tiebreaker, but only if we merely mean a criterion that can be used to select between generally otherwise equally good. But on this definition of a "tiebreaker", all criteria that we use to select between competing theories are "tiebreakers". The criterion of simplicity (and thus judgments of beauty)—like all criteria of theory selection—is not definitive. It must be used in tandem with other methods and criteria of theory evaluation. But neither is simplicity (and thus judgment of beauty) an inferior criterion that can only be deployed after all others have been exhausted. It operates on an equal basis alongside all other criteria of theory selection.

But second, even supposing that I granted that the criterion of simplicity (and thus judgments of beauty) plays only a weak tiebreaking role in theory selection, I also argued that judgments of beauty are part of reflective equilibrium. Namely, I argued the judgments of coherence that we must make when using the reflective equilibrium are a species of judgment of beauty. Thus, given that reflective equilibrium is a chief method for theory construction, judgments of beauty are also required during the construction stage of theory evaluation. This goes further than the previous point. Not only are judgments of beauty part of selecting between competing theories (as part of the criterion of simplicity), they are also part of the fundamental method we use to construct good theories in the first place (as part of the method of reflective equilibrium). This means that judgments of beauty—far from being irrelevant to theory evaluation or merely a weak tiebreaker between otherwise equally good theories—are a fundamental part of how we evaluate theories, both in the construction and selection stages of theory evaluation.

We have reached the final conclusions I definitively argue for: judgments of coherence and simplicity are types of aesthetic judgments (that is, judgments of beauty). This gives the methods of reflective equilibrium and simplicity a complete explanation for how to arrive at preferred theories. And this means that these methods of theory evaluation are themselves aesthetic (or partially aesthetic): they instruct us to use judgments of beauty to select better theories. Part of our task in evaluating philosophical theories will thus be developing the skill of aesthetically evaluating theories. I used modest claims about beauty to get us to these conclusions. But in order to help illustrate these conclusions, I briefly explore a candidate full theory of beauty to show how one might expand on the aesthetic nature of judgments of coherence and simplicity (and, in turn,

the methods that invoke these judgments). I explore Mothersill's theory of beauty, but I might have explored other theories. Other theories of beauty include beauty as harmoniousness, beauty as organic unity, and, perhaps the most venerated theory, beauty as unity in diversity.[95]

Section 5: An Illustration via Mothersill's Theory of Beauty

Mothersill provided the precise formulation of the two theses that I used in my analysis of beauty above. I said that these theses only formed a modest foundation of a theory of beauty, not a full theory. But Mothersill does go on to give a complete (or at least a more complete) theory of beauty. This theory, she says, is a working out of a formulation of beauty that Aquinas gives. Aquinas' formula is: "*Pulchrum dicatur id cujus apprehensio ipsa placet*—'Let us call that beautiful of which the apprehension in itself pleases.'"[96] This theory, obviously, reintroduces the notion of pleasure to the aesthetic. (This is plausibly a strength of the theory, given the intuitive connection between the beautiful and the pleasurable as well as the etymological connection between "aesthetic" and the senses.)

The critical notion for this theory of beauty, aside from pleasure, is "*apprehensio ipsa*", "apprehension in itself". This, Mothersill says, is acquaintance knowledge: knowledge of a thing that is not reducible to propositional knowledge and can only be got by "getting to know" the object. We can know a person with acquaintance knowledge by having moderately intimate conversations with them, and we can know an object with acquaintance knowledge by carefully examining the object. (The exact nature of acquaintance knowledge is, of course, itself difficult and the subject of debate.)[97] Now, Mothersill defines those properties of objects that can only be known by acquaintance—the properties that are particular to a particular object, that give it its particularity—as "aesthetic properties".[98] It's worth noting that this is a technical and abstruse definition of "aesthetic properties". On this definition, a property's being aesthetic has nothing yet to do with being beautiful or even with being a source of pleasure. They are only those properties that are particular to

---

[95] See Aristotle *Poetics*, Burke (1765), Collingwood (1945), Hogarth (1753), Hume (1757), Hutcheson (1725), Kant (1790), Plato *Symposium,* Plotinus *Ennead*, Schiller (1985), Shaftesbury (1711), and Sircello (1975) for other (more or less) full theories of beauty.

[96] Mothersill (1984), p. 323. See Aquinas (1952), Ia, IIae, 27.1, Reply to Objection 3.

[97] And it's worth noting that Mothersill is using the common sense notion, not the technical notion from Russell. See Hasan and Fumerton (2014) for an overview of acquaintance knowledge.

[98] "Aesthetic properties", in this sense, are something like "quiddities".

a particular object, the properties of an object that can only be known by acquaintance knowledge.

Mothersill's theory of beauty puts all this together and says: "Any individual is beautiful if and only if it is such as to be a cause of pleasure in virtue of its aesthetic properties."[99] That is, an object's aesthetic properties (the properties particular to that object, which can only be known by acquaintance knowledge) being the cause of pleasure is what makes that object beautiful. For example, one evaluates the beauty of a piece of music by first getting to know by acquaintance the piece of music. Then, second, one sees whether the piece of music causes pleasure upon hearing and is attentive to whether this pleasure is caused by the properties particular to the piece or by something else. (Mothersill's theory thus preserves, in requiring the pleasure to be caused by the particular properties of the object, the "purity" or "disinterestedness" requirement of judgments of beauty that is common to many theories of beauty.) On Mothersill's theory, this same pattern of aesthetic judgment will hold for any object: first, get to know by acquaintance the object, second, see whether the particular properties of the object are the cause of pleasure. This theory, like any theory, comes with its own puzzles and difficulties. It is not my purpose to fully reproduce Mothersill's theory here. I provided only a summary statement of it, Mothersill has much more to say in its defense.[100] Nor am I endorsing Mothersill's theory, I only lay it out to provide an example of how one might further spell out how judgments of coherence and simplicity are types of judgments of beauty.

So, how would one further spell out how judgments of coherence and simplicity are types of judgments of beauty using Mothersill's theory? If judgments of coherence and simplicity are types of judgments of beauty, and if Mothersill's more complete theory of beauty is correct, then a theory's being coherent or simple means that the aesthetic properties of that theory (again, remember that this means only the properties that are particular to the theory that can only be known by acquaintance) are a cause of pleasure. In this case, the philosophical theory is the relevant object that is being known by acquaintance and evaluated for whether its particular properties are causes of pleasure. Obviously, it is an abstract object (akin, perhaps, to a game of chess), but it still has a set of properties that are particular to it.[101] It is important, on Mothersill's theory, that the theory be an object, since it is required for her account that it can be a cause of pleasure. Again, if judgments of coherence and simplicity are types of judgments of beauty, then (on Mothersill's theory) they are judgments that rely on acquaintance knowledge to see whether the particular properties of an object (like a philosophical theory) are the cause of pleasure.

We could differentiate simplicity and coherence from each other and from other species of beauty by mapping them on to aesthetic properties. There are

---

[99] Mothersill (1984), p. 347.
[100] See Mothersill (1984), Chapter 11-12.
[101] See Osborne (1964).

different types of beauty, so there are different types of aesthetic properties that cause pleasure. Thus, on this view, a theory's being coherent is its being a cause of pleasure in virtue of its coherence-exhibiting aesthetic properties. Likewise, a theory's being simple is its being a cause of pleasure in virtue of its simplicity-exhibiting aesthetic properties. Notice that a theory's being coherent (or simple) is not simply a matter of having coherence-exhibiting (or simplicity-exhibiting) aesthetic properties. It is when those aesthetic properties are causes of pleasure that the theory is judged to be coherent of simple. This is because they are types of beauty: on Mothersill's theory, an object is not beautiful in virtue of having aesthetic properties. All objects have aesthetic properties, using Mothersill's definition of aesthetic properties. It is when those aesthetic properties are the cause of pleasure that the object is beautiful, according to Mothersill's theory. Likewise, when a theory's aesthetic properties (the particular properties that only that theory has, which can only be known by acquaintance with the theory) are coherence- or simplicity-exhibiting and are causes of pleasure, then that theory is judged to be coherent or simple, beautiful in the style of coherence or simplicity.

We can go one step further and plug this explanation of coherence and simplicity in to the methods of reflective equilibrium and simplicity. If Mothersill's theory of beauty is correct, then the methods instruct us to choose a theory by becoming acquainted with the theory's aesthetic properties and seeing whether those aesthetic properties cause us pleasure. That is, the piece of the method of reflective equilibrium that calls for a judgment of coherence can be explained as a call to see by acquaintance whether the coherence-exhibiting aesthetic properties of the candidate theory cause pleasure. When such coherence-exhibiting aesthetic properties of the candidate theory do cause pleasure, then the method of reflective equilibrium marks that theory off as preferred (in conjunction with the other mechanics of the method). The criterion of simplicity can, likewise, be explained as a call to see by acquaintance whether the simplicity-exhibiting aesthetic properties of the candidate theory cause pleasure. When such simplicity-exhibiting aesthetic properties of the candidate theory do cause pleasure, then the method of simplicity marks that theory off as preferred.

It will help to think through an example of how this might look. Return again for a final time to Colonel Mustard and Mr. Boddy. Given my aesthetic understanding of reflective equilibrium and simplicity and Mothersill's theory of beauty, when we use reflective equilibrium and simplicity we should evaluate the competing theories of who killed Mr. Boddy on the basis of whether those theories cause us pleasure upon acquaintance. So consider: We have found Mr. Boddy dead in the library. We see Colonel Mustard holding the bloody candlestick. We know that Colonel Mustard stands to inherit the estate. Someone says, "I have a theory: Colonel Mustard killed Mr. Boddy!" We consider this suggestion, seeing how it fits with all the information we have and judging it on its coherence and simplicity. When we hold the theory in our mind's eye, does it cause us pleasure simply in virtue of its aesthetic properties—the

properties that make it itself? It seems to me that in such cases we are pleased: that the theory's "leaping out at us" as the best theory is a kind of (very minor) pleasure. But perhaps, given that this is such a mundane case, the pleasure involved is too minor or routine to be noticed. Now hold in your mind's eye the great theories of the philosophical (or scientific) canon—Kant's theory of morality, Rawls' theory of justice. Do they, merely upon acquaintance, merely in virtue of their structure, cause pleasure? It seems to me that they do, in the same way that reading the score of a symphony or a watching a well-played game of chess causes pleasure. If so, this will mean that they are beautiful according to Mothersill's account. And this means that they will be judged as better theories when we use reflective equilibrium or simplicity to evaluate them, inasmuch as their beauty takes the form of coherence or simplicity.

We should be careful at this point not to take the wrong point: I am not suggesting that we abandon our epistemic aims when it comes to evaluating theories and merely go in for whatever theories give us pleasure. My claim is: if using the methods of reflective equilibrium and simplicity requires making judgments of beauty and if Mothersill's account of beauty, beauty as pleasure caused by mere acquaintance, is correct, then pleasure caused by the theory itself is part of how we evaluate the theory for epistemic goodness. That is, pleasure caused by mere acquaintance is beauty, and using the methods of reflective equilibrium and simplicity involves using beauty as a (partial) indicator of epistemic goodness. It is also worth remembering that these are but two methods of theory evaluation: I have never suggested that these exhaust our methods of theory evaluation, nor do I think that these methods should be used without other, non-aesthetic, methods of theory evaluation. Indeed, in the case of reflective equilibrium, I have only argued that judgments of beauty are one part of the method. Our considered judgments are an important element in using reflective equilibrium and make no reference to judgments of beauty.

Going in for Mothersill's full theory of beauty thus helps fill out my recommendation above to develop the skill of aesthetically evaluating philosophical theories. That we should develop such a skill was an upshot of my conclusion that reflective equilibrium and simplicity must use aesthetic judgment in order to be complete methods. On Mothersill's account, aesthetically evaluating philosophical theories would require gaining acquaintance knowledge about a theory and then seeing, by this acquaintance, whether the particular properties of the theory were causes of pleasure. Both of these steps: getting to know a philosophical theory by acquaintance and judging whether the theory is a cause of pleasure are uncommon practice in evaluating philosophical theories. And yet, if my arguments about coherence and simplicity are correct, and if Mothersill's theory of beauty is correct, and if we wish to use reflective equilibrium or simplicity, then this is part of how we should be evaluating theories: We should, ceteris paribus, prefer theories that, on acquaintance, cause us pleasure. This is because (on Mothersill's theory) such pleasure is the mark of beauty. And reflective equilibrium and simplicity say (on my completed

67

aesthetic understanding of them) that we should use judgments of beauty as part of our evaluation of theories.

Bringing in the notions of acquaintance knowledge, Mothersill's technical definition of aesthetic properties, and pleasure helps fill in the landscape of what it is for something to be beautiful. The same expanded explanations can be afforded to judgments of coherence and simplicity inasmuch as they are types of judgments of beauty. That is, having a full theory of beauty helps further explain the nature of our judgments of coherence and simplicity (and thus further explain the methods that rely on these judgments), given that judgments of coherence and simplicity are types of judgments of beauty. Notice too that Mothersill's theory of beauty maintains respect for both the First and Second Theses: Judgments of beauty are clearly genuine and possible on her account (inasmuch as aesthetic properties can be causes of pleasure). Yet her account makes no reference to principles of beauty. We have no generalizations that yield even somewhat determinate judgments of beauty; we simply must find out by acquaintance whether an object's aesthetic properties will cause pleasure. So too does her theory preserve this character for judgments of coherence and simplicity. Yet it is clear that, despite not offering any principles, Mothersill's theory does offer further explanation about the nature of beauty. So too could it offer further explanation about the nature of coherence and simplicity. This has only been an illustrative exercise to show how judgments of coherence and simplicity, as well as the methods that use these judgments, could be filled out, mutatis mutandis, by candidate full theories of beauty.

Section 6: Concluding the Argument

Developing a full theory of beauty is a monumental task, far beyond my intentions here. But the lack of a definitive full theory of beauty should not preclude us from using the notion of beauty to make progress on other difficult questions. My concern here is not directly with the nature of beauty, but with the methods of philosophical theory evaluation, in particular the methods of reflective equilibrium and simplicity. I used the notion of beauty at a critical point in developing an account of those methods: a judgment critical to the full working of each of the methods is best understood as a type of aesthetic judgment. So, the nature of beauty and the status of two of our methods of theory evaluation have become entangled. This strategy for developing an account of these methods does not depend on having a full theory of beauty, but it does bring the nature of beauty to the center of our attention and reemphasize the need for a more complete understanding of beauty. We are certainly familiar with the need for a more complete understanding of, say, rightness. But beauty is a more neglected piece of the landscape of value. This project is a small way of addressing that neglect.

To repeat then, these are the only conclusions I definitively argue for: judgments of coherence and simplicity are types of aesthetic judgments (that is, judgments of beauty). Identifying them as species of beauty completes the accounts of the methods of reflective equilibrium and simplicity, which were lacking explanations of judgments of coherence and simplicity. This thus advances our understanding of those methods. And this means that these methods of theory evaluation are themselves aesthetic (or partially aesthetic): they instruct us to use judgments of beauty to select preferred theories. This means that beauty is part of what makes theories (theoretically) good, which advances our understanding of theories. Part of our task in evaluating philosophical theories will thus be developing the skill of aesthetically evaluating theories. I used modest claims about beauty to get us to these conclusions.

I purposefully limited this project to examining only two specific methods of theory evaluation. There are many other methods of theory evaluation that are worth careful analysis. But—though it is far beyond the scope of what I argued for since I cannot consider all methods of theory evaluation—I recommend the lesson: "Wherever one sees a method of theory evaluation that invokes genuine yet unprincipled judgment, one has found an aesthetic method of theory evaluation (a method that instructs us to evaluate theories using, at least in part, judgments of beauty). That is, wherever one sees a value (a piece of the good) about which we make genuine judgments, yet to which we cannot give determining criteria (principles that could feature in a deduction), one has found beauty. Any method of theory evaluation that invokes such a value will be (at least partially) aesthetic." Such an inductive projection is a speculation, not a conclusion. This speculation is a fruitful guide to thinking about the ways we evaluate theories.

I argued that understanding coherence and simplicity as species of beauty can complete the accounts of methods of reflective equilibrium and simplicity. But I did not say a word about whether such methods, especially under my aesthetic completion of them, are justified methods of theory evaluation. One might have started with the presumption that they were justified methods of theory evaluation. But, upon realizing that they need judgments of beauty to complete them, one might conclude: "So much the worse for reflective equilibrium and simplicity. I was prepared to accept them, but seeing that they require judgments of beauty is a reductio ad absurdum for them." Or the objection might be put in a slightly different way: "Using judgments of beauty to select theories cannot be justified. Evaluating theories is an epistemic matter, and beauty has nothing to do with epistemic matters. Thus, any method that requires judgments of beauty is thereby an unjustified method of theory evaluation." What might a proponent of reflective equilibrium and simplicity, even on my aesthetic completion of them, say to such objections? How could (at least partially) aesthetic methods of theory evaluation be justified? I take up these questions in the next chapter. Very broadly speaking, my strategy for resolving

these objections will be to use a transcendental argument that concludes that aesthetic methods of theory evaluation can be justified.

CHAPTER FIVE

JUSTIFYING AESTHETIC METHODS OF THEORY EVALUATION

Section 1: Justified Methods

In the previous chapter I argued that coherence and simplicity are species of the genus of aesthetic judgments (that is, judgments of beauty). This species/genus claim gives an explanation of our judgments of coherence and simplicity. This, in turn, completes the explanation of how the methods of reflective equilibrium and simplicity work, how they evaluate theories as better. And this means that these methods of theory evaluation are themselves aesthetic (or partially aesthetic): using these methods requires that we use judgments of beauty to select better theories. One might accept this argument but then turn it into a *modus tollens*. That is, one might leverage reflective equilibrium and simplicity's need for judgments of beauty into an argument that we should reject these methods. Indeed, those who are already set against reflective equilibrium and simplicity might use my conclusions as evidence for the methods' failure. I think both reflective equilibrium and simplicity are good and powerful methods of theory evaluation, though I did not directly argue for this, so I do not recommend this line of thought. By way of concluding my aesthetic understanding of the methods of reflective equilibrium and simplicity it will be fruitful to consider how one might argue that (partially) aesthetic methods of theory evaluation (such as reflective equilibrium and simplicity) could be justified.

For a method of theory evaluation to be justified means, roughly, that using that method results in theories that deliver genuine understanding of the relevant subject matter.[102] Recall that we began this enterprise by asking, in service to a theory of theories, for an account of what makes theories better. And remember that the relevant notion of "better theories" has never been "better in any way whatever", but rather "theoretically better, better at fulfilling the purpose of a theory." As we discussed in Chapter One, the purpose of theories is ultimately to get us to understanding. This should be the final ground of what makes a theory theoretically better. But this doesn't tell us much about what makes a theory theoretically better, since we can simply push the question back one step and ask "What makes a theory a source of genuine understanding?" Thus, in order to make some progress on the question of what makes a theory better we examined two candidate methods of theory evaluation. Candidate methods of theory evaluation purport to have a method for identifying better

---

[102] It's worth emphasizing that the objects of the justification in question are *methods*. This is distinct from the notion of justification as applied to *beliefs*.

theories, theories that deliver understanding. But candidate methods need to defend themselves: to give an account of how and why they have a method for getting to theories that deliver understanding. This is the question of justification of methods: To find that a method is justified is to promote it from a candidate method of theory evaluation to a justified method of theory evaluation. We explored defenses of reflective equilibrium and simplicity in their earlier respective chapters. But now the danger of reflective equilibrium and simplicity being unjustified methods of theory evaluation (methods that do not actually get us to theories that deliver understanding) has returned with a new face: judgments of beauty. To find that the methods require judgments of beauty might be tantamount to finding that the methods are unjustified. This, then, is the final question I will take up: Could methods that use judgments of beauty to evaluate theories (as I have argued is the case for reflective equilibrium and simplicity) be justified methods?

In one sense, however, I am not committed to resolving this question. To argue for the dilemma "Either reflective equilibrium and simplicity are (partially) aesthetic methods of theory evaluation or they are unjustified methods of theory evaluation" is enough. This is because concluding that reflective equilibrium and simplicity are failures as methods of theory evaluation has such a high theoretical cost. These methods enjoy widespread usage and a long pedigree in philosophy. To find, for instance, that we should not prefer simpler theories would be a terrible blow to our project of finding good philosophical theories. Without reflective equilibrium or simplicity, how will we settle on the theory that Colonel Mustard killed Mr. Boddy? The methods of reflective equilibrium and simplicity are the best methods we have for organizing our judgments (or the scientific analogue, data) into systematic theories. Without them, we'll be left merely with our judgments or the data: We'll know that Mr. Boddy was found dead in the library and that Colonel Mustard was holding a bloody candlestick, but we won't be able to move from this data to a systematic explanation. Thus, resolving the dilemma by accepting that reflective equilibrium and simplicity are (partially) aesthetic methods of theory evaluation has a far lower cost than losing those methods. To be driven to reject powerful and popular methods of theory evaluation because of a worry about beauty reveals a dogmatic opposition to judgments of beauty. Nevertheless, one might have independent worries about reflective equilibrium and simplicity that combine with this dilemma to resolve against these methods. Or one might be able to raise problems with using judgments of beauty for theory evaluation in a non-dogmatic way. But I would be happy to see such debates unfold. To discuss reflective equilibrium and simplicity in terms of this dilemma is a step in the right direction.

Section 2: Justifying Reflective Equilibrium and Simplicity

Suppose that one didn't have a dogmatic opposition to using judgments of beauty in methods of theory evaluation and agreed with my conclusions in the previous chapter. Nevertheless, one might still have questions about why aesthetic judgment can be justifiably used in evaluating theories or why aesthetic judgment might be a good guide to better theories. To answer these questions would take some of the pain out of accepting the first horn of the dilemma: that reflective equilibrium and simplicity are (partially) aesthetic methods of theory evaluation.

The most tempting way to raise a problem with using judgments of beauty to select for theories is to object: "But beauty is, as they say, 'in the eye of the beholder'. How could a quality like that guide us to better theories? Being a better theory isn't merely in the eye of the beholder, so there cannot be a connection between a theory's being beautiful and its being a better theory." But it is difficult to state precisely what this adage means or why it might generate a problem for using judgments of beauty as a guide to better theories. There are two ways that we might understand the adage, but neither understanding generates a problem.

First, we might understand the adage to assert thoroughgoing relativism about beauty. A value is thoroughly relative, roughly speaking, if there is no interesting overlap between individuals' relevant value judgments and no adjudication between such judgments. Then the objection is that such a thoroughly relative value cannot be a guide to better theories. This is because being a better theory is not supposed to be such a thoroughly relative value. Although theories of beauty are as diverse as any subject in philosophy, philosophers of beauty have almost universally agreed that beauty is not relative in this way.[103] Thoroughgoing relativism about beauty is also contrary to common sense: there is widespread agreement that some artworks are more beautiful than others, that some natural environments are more worthy of preservation because of their beauty, etc. The claim "Beauty is relative to each individual human person such that there is no interesting overlap between individuals' judgments of beauty and no adjudication between such judgments"— the claim that there is thoroughgoing relativism about beauty—can be justifiably

---

[103] I say "relative *in this way*" because it is important that the adage expresses, on this understanding, *thoroughgoing* relativism about beauty. It would be a more reasonable position to assert that beauty is relative, say, to the human species. But such relativism to the human species is not what the adage asserts. Nor would such relativism to the human species raise a problem for using judgments of beauty to select for better theories. *Thoroughgoing* relativism about beauty, nevertheless, seems to many (including many contemporary philosophers) to be unquestionable common sense, despite the fact that it is the extreme minority position in the history of philosophy. Thoroughgoing relativism about beauty is given its best philosophical defense in Santayana (1896).

dismissed out of hand. The situation is akin to the situation in ethics: Though the history of ethics is filled with disagreement about morality, there is almost universal agreement that morality is not thoroughly relative. The claim that it is relative to each individual person completely upends our moral judgments and practices.[104] The same is true with respect to our aesthetic judgments and practices. Such an upending of our judgments and practices is enough for a reductio ad absurdum for the relativism in question.

There are, of course, difficult questions about where exactly the value of beauty is located: Is it a property that, strictly speaking, is located in external objects? Or is it located, strictly speaking, in agents' minds? There are also related, but distinct, questions about whether the existence of beauty depends on the existence of minds. But such questions about the metaphysical location and status of the value of beauty are distinct from the relativism question. Rejecting thoroughgoing relativism about beauty need not commit one to any particular answer to the above questions. I am thus not committed to any particular answer to the above questions and I do not take them up here.

Second, we might understand "beauty is in the eye of the beholder" to assert pluralism about beauty. A value is plural if there are multiple, but conflicting, ways that the value can be instantiated. Pluralism about a value is importantly different from thoroughgoing relativism about a value. A value's being plural does not mean that anything can instantiate that value (which does follow from a value's being thoroughly relative). That is, a plural value still has limits on how it can be instantiated. Beauty is plausibly a plural value. For example, there are multiple, but conflicting, ways in which a painting can be beautiful. The canons and methods of realist landscape as compared to abstract impressionism are clearly in conflict. But particular paintings in either school can be beautiful. Given such pluralism about beauty, it is perfectly natural that different instantiations of beauty will appeal to different people (which is plausibly what the adage is trying to communicate). The situation with beauty is analogous, for example, to the notion of the "good life". Not any form of life can qualify as good (thus, the good life is not a thoroughly relative notion). But there are multiple, conflicting ways in which a life can be good. For example, a life as a parent and a life as a monk are both plausibly forms of the good life, yet they are in conflict. The fact that the good life is plural does not undermine our judgments about what is and is not a good life. So too with the case of beauty.

Thus, though beauty is plausibly plural (and so this understanding of the adage is plausibly true), this is no threat to the claim that judgments of beauty can guide us to better theories. Judgments of a plural value can still delimit the range of phenomenon in the way needed. That is, beauty's being plural does not mean that there are no limits on how theories can be beautiful, nor does it mean that any theory whatever can be beautiful. It simply means that there are

---

[104] Though the question of relativism about morality is distinct from the question of moral realism, the two are often discussed in tandem. See Cuneo (2011), Foot (1972), and Joyce (2015).

multiple, conflicting ways in which a theory can be beautiful. This sounds plausible, but it still allows that judgments of beauty can be part of how we delimit good theories.

Thus, if we understand "beauty is in the eye of the beholder" as asserting thoroughgoing relativism about beauty, then the adage is false. But if we understand the adage as asserting pluralism about beauty, then there is no problem. Either way, the adage does not generate an objection to using judgments of beauty to guide us to better theories. So, if citing "beauty is in the eye of the beholder" is not the appropriate way to raise worries about using judgments of beauty to select for better theories, then what is the appropriate way? One might accept that beauty is plural and is not thoroughly relative yet still worry that judgments of beauty cannot be justifiably used to select for better theories. This is because, roughly speaking, there seems to be no reason for beautiful theories to coincide with good theories. Remember that the relevant notion of "better theories" is "theoretically better, better at being a theory, better at delivering understanding." This is an epistemic notion.   You might think that there is a category mistake or an equivocation on "better" or "good" to maintain that judgments of beauty can be justifiably used to select for theoretically better theories. A theory's being better is related to the understanding that it delivers, and a theory's beauty does not seem connected to such an epistemic value.

We could, alternatively, frame the worry this way: The judgments that we use to evaluate better and worse theories should be cognitive judgments, not affective judgments. This is because the evaluation of theories—the construction and selection of theories—is a cognitive matter, not merely a matter of our affects. Judgments of beauty do seem to take the form of cognitive judgments: from a purely grammatical standpoint, they make a judgment about the world, they don't express an emotion. "The painting is beautiful, the landscape is beautiful, the theory is beautiful." These judgments don't, at least not in a wooden literal way, say anything about the speaker's affects. But despite this, many people (including many contemporary philosophers) are inclined to think that judgments of beauty actually express an affective judgment, not a cognitive judgment.[105] That is, they think that the judgments express something about the speaker, not the world. Why might one think this? Perhaps the best reason comes from what we've already discovered about judgments of beauty: they cannot be explicated using principles. How could a judgment that has no explication in terms of principles be cognitive? If a judgment cannot be explicated—and judgments are explicated by reference to principles—then the judgment seems like it is a merely private judgment, an affective judgment, rather

---

[105] See Santayana (1896) for the best defense of the view that judgments of beauty are actually affective, despite the appearance to the contrary. This view—emotivism—is obviously a close cousin to emotivism about ethics, which maintains the analogue about ethical judgments: that ethical judgments are affective judgments, despite their not taking that grammatical form. Both emotivism about ethics and emotivism about ethics have their roots in logical positivism.

than a judgment about the world. But if it is true that judgments of beauty are private and affective, then they would be inappropriate for use in evaluating theories, given that the evaluation of theories is a cognitive matter.

To repeat, then, the objection is that there is no connection between beauty and understanding, that beauty is affective and so disconnected from the cognitive evaluation of theories. But there are ways to vindicate a connection between a theory's being beautiful and its being a source of understanding, which would also indicate that judgments of beauty are cognitive. I said earlier that my argument in the previous chapter (that using reflective equilibrium and simplicity requires making judgments of beauty) could be used to establish a *modus tollens* against those methods. But if that is true, then the argument can also be used to establish a *modus ponens* that concludes that a theory's beauty is related to a theory's being a source of understanding. This argument would go something like: "Reflective equilibrium and simplicity are methods that can deliver better theories. Reflective equilibrium and simplicity evaluate theories in part based on whether the candidate theory is beautiful (specifically, whether it is coherent or simple). Therefore, in the context of using reflective equilibrium and simplicity, a theory's being beautiful is connected to its being a better theory (that is, being able to deliver understanding)."

This *modus ponens* will be as strong as the methods themselves. This strategy for establishing a connection between a theory's beauty and its goodness is something like: "Take our best methods of choosing better theories, see what values those methods connect being a better theory with, and then accept that there is such a connection between those values and being a better theory."   Methods of theory evaluation will always involve such a mediated strategy. We don't have a direct way of evaluating the theoretical goodness of a theory, thus we must connect a theory's goodness with some other feature of a theory that we can directly evaluate. Thus, as long as we depend on methods to get to better theories (as opposed to being able to directly evaluate the goodness of a theory), we will be depending on some connection between a theory's goodness and some other value.

This argument can be fleshed out in two different ways. Which version you prefer will depend on your philosophical predilections. First, the argument can be fleshed out in a Kantian mode by means of a transcendental argument. A transcendental argument, from the work and tradition of Kant, is an argument that takes roughly the following form: "X is a necessary condition for the possibility of Y, where Y is a fundamental condition of human experience. Y is the case. Therefore, X is the case." Applying this argument form to the matter at hand would yield: "A connection between a theory's beauty and its goodness (its being a source of understanding) is a necessary condition for the possibility of reflective equilibrium and simplicity—which are fundamental theoretical methods that humans use—getting us to better theories. Reflective equilibrium and simplicity do get us to better theories (that is, reflective equilibrium and simplicity are justified methods of theory evaluation). Therefore, a theory's beauty is

76

connected to its goodness (its being a source of understanding)." Formalizing the argument in this way does not excise the dependence on the strength of the methods. It still depends on the assumption that reflective equilibrium and simplicity will get us to better theories. Of course, this isn't to say that this is a groundless assumption: We examined the methods in detail in their respective earlier chapters and their ambitions to be justified methods of theory evaluation. But, again, this argument does give out somewhere: It depends on, and does not itself argue for, the strength of the methods themselves. They are the feature of experience that we take as a given in the transcendental argument ("Y"). I have argued that a theory's beauty being connected to its theoretical goodness ("X") is a necessary feature of the methods getting us to better theories (because using the methods means looking to theories' beauty to tell which theories are better). This does not argue against skepticism about the methods so much as it maintains that skepticism about the methods has a higher cost than accepting that a theory's beauty is connected its goodness.

Second, the argument can be fleshed out in the mode of a scientific realist. In this mode, the *modus ponens* given above (which concludes that, in the case of reflective equilibrium and simplicity, a theory's beauty is connected to its goodness) is not fortified by a transcendental argument. Rather, the *modus ponens* is fortified by an analogy to perception. Perception is, loosely speaking, a method for evaluating scientific theories. Like reflective equilibrium and simplicity, evaluating good theories by means of perception involves a mediated strategy: it connects good theories with a directly evaluable feature of those theories. In the case of perception, the directly evaluable feature is directly evaluable par excellence: it is perceptible. It is not, of course, that we directly perceive good theories, but that we can directly perceive relevant observations and we evaluate theories in light of those observations. But it is clear in the case of perception that this mediated strategy for picking good theories raises no justificatory concerns. Perception's function as representing the world is so central to our epistemic endeavors that the justificatory question cannot even find purchase. Thus, skepticism about perception is not judged to be more costly than accepting that perception is justified. Rather, the burden of proof is shifted such that skepticism about perception is not the default position and needs special motivation.

The argument that reflective equilibrium and simplicity are justified methods, even though they require aesthetic judgment, can proceed in an analogous way. Using reflective equilibrium and simplicity to evaluate theories is so central to our pursuit of understanding—they are such pervasive and powerful methods for theorizing—that questioning their justification is not the default position. That is, skepticism about those methods requires special motivation. I argued that these methods require aesthetic judgment, so they can transfer their justificatory presumption to such aesthetic judgments. Like the argument in the Kantian mode above, this argument obviously still depends on the strength of the methods. It does not itself argue for the strength of those methods. But, like

perception, the strength of the methods is so great that it carries with it the justification of the methods. Unlike the argument in the Kantian mode, this does not argue against skepticism about using aesthetic judgment as part of reflective equilibrium or simplicity. Rather, skepticism about these (partially) aesthetic methods is shown to require special motivation.

These modes of argument are closely related, but different. They both depend on the strength of the methods in question to transfer justification to the aesthetic judgments that are necessary for those methods. But they take a different tack with respect to skepticism about those methods, and so resolve the justificatory question in different ways. Whichever mode of argument one prefers, the conclusion is the same: Judgments of beauty are a justified part of using reflective equilibrium and simplicity to evaluate theories. That is, in the context of reflective equilibrium and simplicity, there is a connection between a theory's beauty and its being a source of understanding.

And what of judgments of beauty being affective or cognitive? The above arguments can also be framed to give reason to think that judgments of beauty are indeed cognitive, and so can be appropriately used to evaluate theories. The argument can run: "Take our best methods of choosing better theories, see what values those methods connect being a better theory with, and then accept that judgments of such values are cognitive, given that choosing better theories is a cognitive matter." Again, this argument will be as strong as the methods themselves. We can frame this arguments as shifting the burden back on to the emotivist about beauty: One might have thought that judgments of beauty are affective judgments, but our practice and history of using reflective equilibrium and simplicity—methods that use judgments of beauty to evaluate theories— shows that they are indeed cognitive judgments. The emotivist asked how judgments of beauty can be cognitive if they cannot be given principled explication. But now we can shift the burden back on to the emotivist by asking, "Why assume that all cognitive judgments can be explicated with principles? That is, why assume that all explication of judgments will involve principles? Indeed, our practice of using judgments of beauty— significantly including our using judgments of beauty to evaluate theories when we use reflective equilibrium and simplicity— depends on judgments of beauty being cognitive."

Although we cannot explicate our judgments of beauty about theories by reference to principles, we can give reasons for our judgments of beauty.   We give reasons for our judgments of beauty about artworks by pointing out features of the artwork in question: "See how the artist uses color there, that's really beautiful. Listen to how the musical passage develops over time, that's really beautiful." We can, likewise, give reasons for our judgments of beauty about theories: "The two principles don't really fit together, so the theory isn't satisfying. The theory gives the right judgments only by tacking on several overly-complicated exceptions, so it isn't satisfying." Such reason-giving isn't principled, it's merely pointing out features of the object in question. But it is indicative of the judgments' being cognitive (or at least it establishes that the burden of proof

is on the emotivist to show that they are not cognitive). The upshot of discovering that the methods of reflective equilibrium and simplicity require judgments of beauty is not that those methods are thereby unjustified, but that judgments of beauty are part of our cognitive life, part of our rational evaluation of theories. Judgments of beauty can be responsive to reasons, and so appropriate for the evaluation of theories, without being principled.

So, I have concluded that judgments of beauty are indeed cognitive, that there is a connection between a theory's being beautiful and its being a source of understanding. These conclusions are evidenced by the role that judgments of beauty play in our practice of evaluating theories—our use of reflective equilibrium and simplicity. There are several important clarifications of these conclusions that are worth noting. I have used the somewhat loose formulation "a theory's beauty is connected to its theoretical goodness."[106] This looseness is intentional and important. The claim is certainly not as strong as "All beautiful things are also theoretically good." Aside from being implausible, such a claim would generate significant category mistakes. Many things that can be beautiful simply are not the sorts of things that can be theoretically good: sunsets, for example. The claim is limited to the case of theories, the type of object of interest to us here. Theories are not subject to such a category mistake: they can be both beautiful and theoretically good. But neither is the claim as strong as "All beautiful theories are also theoretically good." I am certainly not denying the possibility that a theory could beautiful, but bad (that is, not a source of understanding). My purpose here has been to examine particular methods of theory evaluation. A method of theory evaluation can be justified without having to be infallible. That is, a method need not infallibly guarantee that the theory it settles on is good in order to be a justified method. For a method of theory evaluation to be a justified method of theory evaluation it is enough that competent use of the method sufficiently results in theories that deliver understanding. Thus, the claim "a theory's beauty is connected to its goodness" should be understood as "there is enough overlap between beautiful theories and good theories to justify methods that look to a theory's beauty to help judge its goodness."

Note also that I said that it is enough that a theory's beauty need only help judge the theory's goodness. I have only been examining two particular methods of theory evaluation. This is compatible with using these methods of theory evaluation in conjunction with other methods of theory evaluation, which might have nothing to do with beauty. When using multiple methods of theory evaluation, it is perfectly possible that the non-aesthetic methods of theory evaluation outweigh (or otherwise defeat or supersede) the aesthetic methods of theory evaluation (which judge a theory's goodness by its beauty).

Indeed, I have argued that one of our methods of theory evaluation— reflective equilibrium—is only a partially aesthetic method of theory evaluation.

---

[106] Remember, "theoretical goodness" means "the goodness of a theory, a theory's being a source of understanding".

That is, the judgment of beauty (coherence or fitness) that I argued was necessary for reflective equilibrium to be complete is only one part of how reflective equilibrium is supposed to work. Considered judgments are also integral to reflective equilibrium. This mean that not only could reflective equilibrium be combined with other, non-aesthetic methods of theory evaluation, but it also has non-aesthetic elements internal to the method. It is perfectly possible that these non-aesthetic elements could outweigh the method's judgment of beauty in its final recommendation on what theory is to be preferred. For example, suppose you are comparing two competing theories using reflective equilibrium. Using the methods means judging each theory on how well it fits with the various relevant considered judgments. I argued that this notion of fitness (or coherence) should be understood as a species of beauty and that thus the method is calling for a judgment of beauty, for evaluating the candidate theories with respect to their beauty. The two competing theories might exhibit different degrees of fitness (different degrees of beauty) and using the method does mean, all else being equal, preferring the one that exhibits greater fitness (beauty). But this is not the only way the theories could differ in a reflective equilibrium. They might capture different considered judgments or a different number of considered judgments. The theory that, on the whole, exhibits greater coherence might very well capture fewer or less important of our considered judgments and therefore not ultimately be recommended in the final judgment when using method. On such a scenario, the less coherent (less beautiful) theory would come out as preferred. Thus, even within my aesthetic interpretation of reflective equilibrium, a theory's beauty is by no means definitive.

Though my claim that a theory's beauty is connected to its theoretical goodness should be understood as moderated in the ways examined above, it remains a striking claim. A connection between aesthetic and epistemic values is somewhat mysterious. These conclusions, however, are strictly only meant to apply in the case of reflective equilibrium and simplicity. For all I have said, other methods of theory evaluation could work completely differently. I do not, however, think that the connection between a theory's beauty and its theoretical goodness that I have argued for is peculiar to the two methods under consideration. It is beyond the scope of this project to definitively argue for this, but I will sketch a way forward.

Section 3: Sketching a Way Forward to the Necessity of Aesthetic Judgment for Theorizing

Recall from Chapter One that theories are defined as "systems of explanation". Because theories are necessarily systematic, methods of theory evaluation will often be tuned in to what makes systems better and worse. Now,

what makes a system better or worse? That will largely depend on what kind of system is in question and what extrinsic purpose the system is being put to. That is, we evaluate economic systems by their economic effects, political systems by their justice, explanatory systems (theories) by the understanding they provide, etc. But we can also simply evaluate systems for their intrinsic properties. That is, we can simply evaluate systems as systems (though it may be uncommon that we make such a rarefied judgment). I suggest that when we evaluate systems themselves as better or worse, we evaluate them by their beauty. That is, if we are asked to evaluate a system as better or worse at being a system (regardless of the particular purpose to which the system is being put to use), then we will make a judgment of beauty.

Recall that I argued that the distinguishing mark of judgments of beauty is being unprincipled and yet genuine. We do genuinely judge systems themselves to be better or worse, and yet there do not seem to be any principles to guide us. We can follow Mothersill's strategy for establishing a lack of principles: Try to articulate principles by which we evaluate, say, pieces of music, chess games, or mathematical equations. (These examples are plausible candidates for being relatively "pure" systems, systems that have a relatively thin extrinsic purpose to which they are directed. Yet even in the case of these examples we do not evaluate them as better or worse merely as systems. We also evaluate them in terms of the particular uses to which they are applied. Nevertheless, they are good examples of coming close to evaluating systems merely as systems.) Any candidate principle for evaluating systems themselves will (I predict) be either circular or subject to rampant counterexample. Instead of using principles, we evaluate systems themselves as elegant and the like: "This system is better than that system because it is more elegant." This indicates that judgments of beauty are the way in which we evaluate systems themselves: we evaluate systems as systems as better or worse based on their beauty.

To see this, it will be helpful to examine some examples in more detail. Consider chess games. We do often compare the quality of chess games: "That was a really good game, better than the last one we played." "This particular game shows the master at her best. Notice what she did here, here, and here. This an extraordinary game." What metric are we evaluating such chess games by?   There is, of course, a very thin extrinsic purpose to which chess games are directed, namely, winning chess games. But, as any good sportsman knows, the achievement of the objective of winning does not exhaust the value of the game. If you play chess merely to win and not to enjoy the structure of the game, then you are missing the main point of the game. Our judgments bear this out: "He lost, but he played an excellent game." Perhaps we are evaluating chess games by the skill displayed by the player. That is surely right, but the skill in doing what? In winning chess games? But this has the same problem I just raised: the value of chess (and thus the value of the relevant skill) is not exhausted by mere victory. Rather, the skill we admire in players is simply playing chess well. This skill is foremost about developing excellent systems, strategies, and structures

for playing the game. Successfully doing this will, of course, often result in winning the game, but that is not the main point. The chief value of chess is rather in developing an excellent chess game, a system of play. When two masters play against each other—each with their own excellent systems of play—then they are forced to adapt their systems, probe their opponent's systems, and execute their systems in increasingly complex circumstances. Such games are truly superb chess games.

Our judgments in which we contrast the value of chess with the value of winning chess bear this out. Consider the following judgments: "He won, but he shouldn't have; she played better." "Although it ended in stalemate, the game was a dazzling display!" "She won, but she usually plays better." "He's very good at winning, but he relies too much on cheap tricks." Although the rules and objectives of chess give it a necessary structure, the value of chess is not simply to game those rules so that one can win every time. The first of these judgments—"He won, but he shouldn't have; she played better"—is the most striking. If playing the best game of chess merely meant winning, this judgment would be nonsensical. The player who wins would, by definition, have played the better game. But we recognize that there is a value in the game—in the structures of play the players are required to develop—that only reliably tracks actually winning the game, it is not identical to winning the game. In well-designed games, playing the better game will very often result in winning. But no game establishes this correlation perfectly. Instead, the value is in simply playing well, where this means developing an excellent system of play and which is theoretically independent from actually winning the game.

We see this in other games as well. Consider the distinction between errors and hits in baseball. Despite having the same gameplay result (the batter or baserunner advancing one or more bases), this distinction reflects a difference in the quality of the game being played. A game with many errors is a worse game than a game with many hits, all else being equal, despite the fact that the score could be the same. If we were only interested in tallying bases or wins, there would be no point in recording this difference. But this distinction is important when it comes to evaluating the quality of players and games.

So, it is clear that we are not merely interested in winning when we make comparative or non-comparative evaluations of games. We are instead evaluating the quality of the game itself: the quality of the structure of the game being played. Good players are not merely those who win often, but those who develop excellent systems of play. Now, what is it to evaluate a system of play, the structure of a game? I suggest that such evaluations are judgments of beauty. That is, judgments like, "He played an excellent game, despite losing," express an aesthetic judgment. Why think this? First, we often express such judgments using explicitly aesthetic vocabulary: "Although his opponent won, he played the more elegant game of chess" "She won, but she had to use that crude strategy to do it."

Second, these judgments have exactly the character that I identified (following Kant and Mothersill) as the special character of judgments of beauty: they are unprincipled, and yet genuine. That the relevant judgments are unprincipled follows from distinguishing the value of a good game from the value of winning. If the value of the game were principled, we could build those principles into the rules of the game such that those who abided by them would always win. But there are no such principles that one could follow in playing, say, an excellent chess game. Despite this lack of principles, we continue to make judgments about the quality of games. We are able to make all-things-considered judgments about the excellence of games without ever invoking a principle. This shows that these judgments about the excellence of games are judgments of beauty. This is analogous to our judgments of beauty about artworks. Traditions of art do often have canons and rules of thumb to follow when producing an artwork (these are the rough analogue of the rules of a game). But these are not to be understood as principles, which if followed would always result in a beautiful artwork. There simply are no principles one can follow to always produce a beautiful artwork, just as there are no principles one can follow to always play a good game of chess.

The purity of such judgments—the fact that such judgments are judgments about the quality of the system without regard to any purpose extrinsic to the system—is what makes them judgments of beauty. If the judgments were to be directed at some extrinsic purpose, we could use principles that directed us to that purpose to make the judgments. But because we are evaluating systems as systems, without regard to the systems' extrinsic purposes, the judgments will be unprincipled and hence judgments of beauty. This is because there are no principles for evaluating the goodness of systems as systems. Meditating on how we evaluate relatively "pure" systems—systems with thin extrinsic purposes, like games and artworks—illuminates this. And if this is right, then it will be true for all systems, inasmuch as we evaluate them as systems. When we evaluate a system as a system we will make a judgment of beauty. Because there are no principles for evaluating the goodness of systems as systems, judgments of beauty are the way in which we evaluate systems as systems. Of course, as I said above, many systems have important extrinsic purposes that they are directed toward (economic systems, social systems, etc.). We will evaluate such systems by how well they serve their particular purpose. But we will also still evaluate these systems as systems—better or worse at being a system, like we evaluate chess games as better or worse at being excellent systems of play. And this evaluation will be a judgment of beauty.

Now return to the case of theories. In the case of theories (systematic explanations), we do not merely evaluate them as systems (theories are not as "pure", in the sense defined above, as the systems that I used as examples). We want theories to be sources of understanding, and so evaluate them using methods we have for detecting theoretical goodness. We also evaluate theories in terms of their compatibility with other good theories in nearby domains. But

this does not escape the fact that theories are systems, and when we are judging theories as systems we will still be making a judgment of beauty. This will be but one part of our evaluations of theories, but an inescapable part as long as theories are systems. Thus, if theories are systems (and they are by definition), then there is no hope of escaping judgments of beauty as part of our evaluation of theories.

Put the point another way: Recall what I said above about methods of theory evaluation always being mediated. They use some middle term to get us to evaluating the theoretical goodness of theories: "Theories that exhibit P are better. This theory exhibits P more than that one. Therefore, we should prefer this theory to that one." Given this mediated nature of methods of theory evaluation (because of our limitations with respect to detecting theoretical goodness), and given that theories are by their nature systematic, methods of theory evaluation will often hook on to some feature of systems as their middle term. Such a feature of systems, used as a middle term to get to theoretical goodness, will, I predict, very often be a species of beauty. This is because systems as systems are evaluated by their beauty.

Now, recall from Chapter One that the systematicity of theories can be connected with the systematicity of understanding. If one is worried about the connection between systems and beauty that I sketched above, yet one still wants to get at epistemic value, one might prefer to go in for unsystematic packages of epistemic value. Such a thing would no longer be a theory, nor would it explain, nor would it deliver understanding, since all these are systematic by definition. But one might set one's epistemic sights lower, and be merely content with truth without understanding. This would keep judgments of beauty separated from our epistemic endeavors. Indeed, without the constraints of systematicity, it might be possible to have an epistemic collection with even more truth: all one has to do is list the various truths that one finds. And yet, despite this possibility, philosophers (and scientists) consistently prefer systematic explanations that deliver understanding to unsystematic lists of truths. This is because we are not merely interested in truth, but in achieving understanding. But this drives us back to systematicity and its connection with beauty. Thus, systematic explanations (theories) will always be evaluable both in terms of epistemic values (whether they deliver understanding) and in terms of beauty. They will be evaluable in terms of epistemic values because of their epistemic purpose, and they will be evaluable in terms of beauty because they are systems.

Now, a fundamental assumption of our theorizing is that pursuing systematic explanations is not a futile endeavor. That is, in our pursuit of theories as answers to philosophical questions, we assume that the right theory could provide us with understanding of the relevant subject. But, given that systems are evaluated as systems with judgments of beauty, we will also build and evaluate these systems—meant to get us to understanding—with an eye to the systems' beauty. That is, in assuming that systematic explanations can get

us understanding, we are assuming that such understanding can be got by way of beauty. Thus, it seems to be an implicit assumption in our pursuit of theories as answers to philosophical questions that there is some connection between beautiful theories and theories that provide understanding.

If this speculative line of thought is right, then we can expand the arguments about the connection between beauty and theoretical goodness to apply to not merely the two methods under consideration, but to theorizing—systematic explaining—itself. First, we can proceed in the Kantian mode by means of a transcendental argument. Recall that transcendental arguments take the form, "X is a necessary condition for the possibility of Y, where Y is a fundamental condition of human experience. Y is the case. Therefore, X is the case." If my speculation about judgments of beauty being the way we evaluate systems as systems is correct, then we can render the argument: "A connection between a theory's beauty and its goodness is a necessary condition for the possibility of systematic explanations (theories)—which are the fundamental way in which humans understand things—providing genuine understanding. Systematic explanations can provide genuine understanding. Therefore, a theory's beauty is connected to its goodness." This argument, like all transcendental arguments, depends on the strength of the cited feature of experience ("Y"). In this case, the feature of experience is the practice of theorizing, the preference for systematic explanations over unsystematic collections of epistemic value. The argument does not itself vindicate our preference for systematic explanations. If we wish to be completely free from using judgments of beauty to select explanations, and if systems as systems are evaluated by their beauty, then we may wish to rid ourselves of our need for systematic explanations. This does not refute skepticism about a connection between a theory's beauty and its goodness so much as it maintains that skepticism about such a connection has a higher cost than accepting the connection. If we assume that we should pursue theories in our pursuit of epistemic value (remember, we could try to be content with mere lists of truths), and if judgments of beauty are a part of how we evaluate theories (because they are systems), then we are implicitly assuming that beauty has some connection to understanding. We can thus frame "a theory's beauty is connected to its goodness" as something of a practical postulate of our theorizing.

Second, we can proceed in the scientific realist mode. Remember that this argument for a connection between a theory's beauty and its goodness (and thus for the justification of using judgments of beauty in theory evaluation) proceeds by analogy to perception. The skepticism about perception is not the default position and requires special motivation because the connection between the perceptible (relevant observations) and theoretical goodness is so central to our epistemic endeavors. If my speculation about judgments of beauty being the way we evaluate systems as systems is correct, then we can expand this realist argument beyond reflective equilibrium and simplicity. Judgments of beauty are part of how we evaluate systems, and our preference for systematic explanations

85

and understanding is so central to our epistemic endeavors that the burden of proof is shifted on to skepticism about using judgments of beauty in our epistemic endeavors. Again, this argument does not vindicate our preference for systematic explanations. But given such a preference, skepticism about a connection between a theory's beauty and its goodness is not the default position and requires special motivation. The justification of using judgments of beauty in evaluating theories has the same realist ground as the justification of using perception in our epistemic endeavors: their function as representing the world.

These lines of thought about expanding the connection between a theory's beauty and its goodness are speculative and inchoate. The conclusions I wish to argue for are strictly limited to the case of reflective equilibrium and simplicity. To expand my conclusions, it would be fruitful to take up other specific candidate methods of theory evaluation to see if they too require judgments of beauty to be complete. These speculative lines are meant to give hope to such expansions to other methods. It also puts more pressure on anyone who is tempted to use my aesthetic completion of reflective equilibrium and simplicity as a means to reject those methods as unjustified. If my speculation is on the right track, then the problem will not be limited to reflective equilibrium and simplicity. This would mean biting an even larger bullet (and rejecting reflective equilibrium and simplicity is already a large enough bullet) in order to deny the role of aesthetic judgment in choosing better theories.

Section 4: Conclusion

To argue that a theory's beauty is connected to its goodness in the case of reflective equilibrium and simplicity and to speculate that they are more widely connected may still seem like a fool's errand. For most lay people, and for many philosophers, thoroughgoing relativism about beauty is hard to shake. But it's worth remembering that our tradition, the source of "beauty is in the eye of the beholder", also contains another adage: "*Pulchritudo splendor veritatis*—Beauty is the splendor of truth."[107] This is not an argument; it does no good to clash one adage with another. But this does show that the claim that there is a connection between epistemic value and beauty has a long pedigree, and has not always been written off as preposterous. Indeed, perhaps the most extensive writings on a connection between beauty and understanding come from natural scientists of the 20th century. Despite the hard-nosed, empiricist reputation of the natural sciences, these thinkers (particular those around the development of relativity and quantum theory) were running up against the limits of experimentation and were driven to such thoughts. Heisenberg attributes the following to Einstein: "If

---

[107] This adage is traditionally attributed to Plato, and certainly fits with his thought, but he never expresses it in precisely these words.

nature leads us to mathematical forms of great simplicity and beauty—by forms I am referring to coherent systems of hypotheses, axioms, etc.—[…] we cannot help thinking that they are 'true,' that they reveal a genuine feature of nature."[108] And from Penrose: "It is a mysterious thing in fact how something which looks attractive may have a better chance of being true than something which looks ugly."[109]   And from Dirac: "One has a great confidence in the theory [of relativity] from its great beauty, quite independent of its detailed successes."[110]   Again, marshaling such quotes is not an argument. But they do show the pedigree of thinking there is a connection between beauty and epistemic value.

Thinking that there is a connection between beauty and epistemic value also has another pedigree, one closer to home. Although Rawls never speaks explicitly of a connection between beauty and epistemic value, he does at times use aesthetic metaphors to describe his theory of justice. He frames his task of laying out a theory of justice as "painting a picture" that he hopes that we will be drawn to. He says that he hopes that people can come to "see themselves" in his theory of justice, that people will develop an "allegiance" to his theory of justice. This motif is inherited from Rousseau. On Rousseau's account, the chief problem of justice is that people-as-they-are are so deeply unjust that they are incapable of working up a just society. The solution is to have a legislator paint a picture of a just society that people, even as they are, will be drawn to. By means of this aesthetic draw, people can change over time until a just society is reached. This is also reminiscent of the third stage of theory evaluation that I outlined at the very beginning of this dissertation: Kuhnian reorientation. Rawls and Rousseau seem to have something like this in mind when they hope that their theories of justice will be so striking that they are able to help us undergo a paradigm shift, seeing our social world—the object of their theorizing—in a new light. Beauty is a perfectly natural part, indeed an integral part, of such theory evaluation. That is, beauty is at least relevant to epistemic goodness at this most mysterious stage of theory evaluation.   One might have been prepared to accept a connection between beauty and theoretical goodness if it were limited to only this Kuhnian stage of theory evaluation. But I have also argued in this dissertation that this same phenomenon—a connection between beauty and theoretical goodness—is present at all stages of theory evaluation. We have seen it in the construction phase of theory evaluation in my (partially) aesthetic understanding of reflective equilibrium—a chief method for theory construction. And we have seen it in the selection phase of theory evaluation in my aesthetic understanding of simplicity—a chief criterion for theory selection. Although I have not framed it this way throughout this dissertation, we could frame these

---

[108] Heisenberg (1971), p. 68. Note how he speaks of *simplicity, coherence,* and *beauty* all in the same breath.
[109] Penrose (1974), p. 267.
[110] Dirac (1980), p. 44. For further cataloguing and discussion of these and other scientists' remarks on the connection between epistemic value and beauty, see McAllister (1996), particularly Chapter 6.

connections between beauty and theoretical goodness in the same terms as above: Beauty is relevant to theoretical goodness because a theory's beauty draws us to that theory, makes us allegiant to that theory. Such drawing to theories is part of what makes a theory good.

Despite this sweeping claim, I have only argued for a small piece of such a connection between beauty and understanding—limited to reflective equilibrium and simplicity. My argument ran: The methods of theory evaluation of reflective equilibrium and simplicity require, respectively, judgments of coherence and simplicity in order to give us a complete account on how to evaluate theories. Such judgments of coherence and simplicity are best understood as species of judgments of beauty, because they are unprincipled and yet genuine. Thus, the methods of reflective equilibrium and simplicity require judgments of beauty in order to be complete methods. They are best understood as (partially) aesthetic methods of theory evaluation. That is, they are best understood as methods of theory evaluation that judge a theory's goodness (in part) by its beauty. Because of the strength and indispensability of these methods of theory evaluation, this amounts to an argument for a connection between beauty and understanding. That is, if two of our best methods of theory evaluation evaluate theories that deliver understanding (in part) by their beauty, then there is such a connection between understanding and beauty. And this is to say that beauty is part of what makes theories good. Remember, my strategy for answering the question "What makes a theory good?" was to analyze two of our methods for evaluating better and worse theories. Because our practice in philosophy is predicated on the claim that using these methods gets us to better theories, we can discover what makes a theory good by analyzing these methods. And by the analysis of our practice—our use of reflective equilibrium and simplicity and why those methods are justified—we have discovered that beauty is part of what makes theories good.

These conclusions—reflective equilibrium and simplicity are (partially) aesthetic methods, there is a connection between beauty and understanding, beauty is part of what makes theories good—mean that judgments of beauty have a justified place in our theorizing, in our evaluating theories. Such judgments of beauty would not, of course, be definitive; nor does this conclusion denigrate other kinds of non-aesthetic methods and judgments for theory evaluation. But it does mean that, as part of the complete picture of how we evaluate theories, we will need to be attentive to our judgments of beauty. This means we should develop our skill at making such judgments: developing the skill of judging theories' beauty. We would be aided in such an endeavor by refining our aesthetic judgment in other domains and by developing canons of aesthetic judgment for philosophical theories.

This serves to complete the part of the theory of theories that I set out to build. I sketched accounts of the nature and purpose of theories in Chapter One, but only now have we reached an (partial) account of what makes theories better. My aesthetic completion of the methods of reflective equilibrium and

simplicity means that beauty is an important part of what makes theories better, even theoretically better. My strategy for making progress on the question "What makes a theory better?" was to examine the structure and justification of two candidate methods. These methods purport to identify better theories—theories that provide understanding—and so they can provide a partial account of what makes theories better. The properties they identify as being part of what makes a theory better can be used to give an explanation, to develop a theory, of what makes theories better. Now, I argued that these two methods—two of our most powerful and popular methods—rely on judgments of beauty to render a complete verdict about which theories are better. As I stressed above, this by no means amounts to the claim that judgments of beauty are definitive about what makes a theory better. This is why this only amounts to a partial theory of what makes theories better. Nevertheless, it represents progress in developing a theory of theories: part of what makes theories theoretically better is their beauty. And if theories' being theoretically better is about their providing understanding, then, again, this amounts to an argument that there is such a connection between understanding and beauty.

Such conclusions are striking, but also modest. This does not, after all, utterly overturn our current practice of philosophical theorizing. Indeed, my strategy was to take philosophical practice as it is and from analyzing that practice discover what we can about what makes theories good. If I am right in my aesthetic completion of reflective equilibrium and simplicity, then we have already been using judgments of beauty whenever we use these methods. This is, rather, a call to recognize, justify, and expand upon these aesthetic judgments. Rather than expunging them from our theorizing, we should acknowledge the judgments of beauty that we make in our theorizing and work to ensure that these judgments are as clear-sighted as possible. In this way, beauty—a precious value in its own right and so near to our hearts and psychologies—can be our ally in our pursuit of understanding.

LIST OF REFERENCES

Allison, Henry. 2001. *Kant's Theory of Taste*. Cambridge University Press.

Alston, William. 2005. *Beyond Justification*. Cornell University Press.

Aquinas, Thomas. 1952. *Summa Theologica*. Trans. Fathers of the English Dominican Province. *Encyclopedia Britannica* (Great Books of the Western World, 19).

Aristotle. *Poetics*. Trans. Ingram Bywater. Clarendon Press, 1920.

Arregui, Jorge V., and Pablo Arnau. 1994. "Shaftesbury: Father or Critic of Modern Aesthetics?" *British Journal of Aesthetics* 34:350-362.

Baker, Alan. "Simplicity". *The Stanford Encyclopedia of Philosophy,* Fall 2013 Edition. Ed. Edward N. Zalta. URL = http://plato.stanford.edu/archives/fall2013/entries/simplicity/

Beardsley, Monroe C. 1973. "What is an Aesthetic Quality?" *Theoria* 39:50-70.

Bergstrom, Lars. 1984. "Underdetermination and Realsim". *Erkenntinis* 21:349-65.

Boyd, Richard. 1973. "Realism, Underdetermination and a Causal Theory of Evidence". *Nous* 7:1:1-12.

Brandt, Richard. 1979. *A Theory of the Good and the Right*. Prometheus Books.

Braithwaite, Richard B. 1953. *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science*. Cambridge University Press.

Brink, David. 1987. "Rawlsian Constructivism in Moral Theory". *Canadian Journal of Philosophy* 17:1:71-90.

Bunge, Mario. 1963. *The Myth of Simplicity: Problems of Scientific Philosophy*. Prentice-Hall.

Burke, Edmund. 1756. *A Philosophical Enquiry into the Origin of Our Ideas of the Sublime and Beautiful*. Ed. Adam Philips. Oxford University Press, 2008.

Cavell, Stanley. 1965. "Aesthetic Problems of Modern Philosophy". In *Philosophy in America*. Ed. Max Black. Cornell University Press.

Chandrasekhar, Subrahmanyan. 1987. *Truth and Beauty: Aesthetics and Motivations in Science*. University of Chicago Press.

Collingwood, R. G. 1945. *The Principles of Art*. Clarendon Press.

Cuneo, Terrance. 2011. "Moral Realism". In *The Continuum Companion to Ethics*. Ed. C. Miller. Continuum.

Curtin, Deane W. 1982. *The Aesthetic Dimension of Science*. Philosophical Library.

Daniels, Norman. 1979. "Wide Reflective Equilibrium and Theory Acceptance in Ethics". *The Journal of Philosophy* 76:5:256-282.

Daniels, Norman. 1980a. "On Some Methods of Ethics and Linguistics". *Philosophical Studies* 37:1:21-36.

Daniels, Norman. 1980b. "Reflective Equilibrium and Archimedean Points". *Canadian Journal of Philosophy* 10:1:83-103.

Daniels, Norman. 1983. "Two Approaches to Theory Selection in Ethics". In *Morality, Reason, and Truth*. Eds. D. Copp and D. Zimmerman. Rowman and Littlefield.

Daniels, Norman. 1996. *Justice and Justification*. Cambridge University Press.

Daniels, Norman. "Reflective Equilibrium". *The Stanford Encyclopedia of Philosophy*, Winter 2013 Edition. Ed. Edward N. Zalta. URL = <http://plato.stanford.edu/archives/win2013/entries/reflective-equilibrium/>

Derkse, Wil. 1993. *On Simplicity and Elegance: An Essay in Intellectual History*. Eburon.

DePaul, Michael. 1986. "Reflective Equilibrium and Foundationalism". *American Philosophical Quarterly* 23:1:59-69.

DePaul, Michael. 1987. "Two Conceptions of Coherence Methods in Ethics". *Mind* 96:463-81.

DePaul, Michael. 1993. *Balance and Refinement: Beyond Coherence Methods of Moral Inquiry*. Routledge.

DePaul, Michael and William Ramsey. 1998. *Rethinking Intuition*. Rowman and Littlefield.

Dickie, George. 1974. *Art and the Aesthetic: An Institutional Analysis*. Cornell University Press.

Dickie, George. 2005. "Iron, Leather, and Critical Principles." In *Contemporary Debates in Aesthetics and the Philosophy of Art*. Wiley-Blackwell.

Dirac, P.A.M. 1980. "The Excellence of Einstein's Theory of Gravitation." In *Einstein: The First Hundred Years*. Pergamon Press.

Douven, Igor. "Abduction". *The Stanford Encyclopedia of Philosophy*, Spring 2011 Edition. Ed. Edward N. Zalta. URL = <http://plato.stanford.edu/archives/spr2011/entries/abduction/>

Dummett, M.A.E. 1973. *Frege: The Philosophy of Language*. Duckworth.

Earman, J. 1993. "Underdetermination, Realism and Reason." *Midwest Studies in Philosophy* 18:19-3Go8.

Elgin, Catherine. 1996. *Considered Judgment*. Princeton University Press.

Engler, Gideon. 1990. "Aesthetics in Science and in Art". *British Journal of Aesthetics* 30:24-34.

Fischer, John Martin and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.

Feuer, Lewis S. 1957. "The Principle of Simplicity". *Philosophy of Science* 24:109-122.

Forster, Malcolm, and Elliott Sober. 1994. "How to Tell When Simpler, More Unified, and Less Ad Hoc Theories will Provide More Accurate Predictions". *British Journal for Philosophy of Science* 45:1-35.

Foot, Philipa. 1958. "Moral Arguments". *Mind* 67:268.

Foot, Philipa. 1972. "Morality as a System of Hypothetical Imperatives". *The Philosophical Review* 81:3.

Foot, Philipa. 2001. *Natural Goodness*. Oxford University Press.

Garthoff, Jon. 2010. "Structuring Ends". *Philosophia* 38:4.

Garthoff, Jon. 2015. "The Priority and Posteriority of Right". *Theoria* 81:2.

Gaut, Berys. 1998. "The Ethical Criticism of Art". In *Aesthetics and Ethics: Essays at the Intersection*. Ed. Jerrold Levinson. Cambridge University Press.

Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23.6.

Ginsborg, Hannah. "Kant's Aesthetics and Teleology". *The Stanford Encyclopedia of Philosophy,* Spring 2013 Edition. Ed. Edward N. Zalta. URL = <http://plato.stanford.edu/archives/spr2013/entries/kant-aesthetics/>.

Glymour, Clarke. 1971. "Theoretical Realism and Theoretical Equivalence". *Boston Studies in the Philosophy of Science* 14:267-97.

Goldman, Alan H. 1990. "Aesthetic Qualities and Aesthetic Value". *Journal of Philosophy* 87:23-37.

Goldman, Alan H. 2005. "There Are No Aesthetic Principles." In *Contemporary Debates in Aesthetics and the Philosophy of Art*. Wiley-Blackwell.

Goodman, Nelson. 1955. *Fact, Fiction, & Forecast*. Harvard University Press.

Goodman, Nelson. 1978. "When is Art?" In *Ways of Worldmaking*. Hackett.

Guyer, Paul. 1979. *Kant and the Claims of Taste*. Harvard University Press.

Guyer, Paul. 1993. *Kant and the Experience of Freedom*. Cambridge University Press.

Hasan, Ali and Richard Fumerton. "Knowledge by Acquaintance vs. Description". *The Stanford Encyclopedia of Philosophy,* Spring 2014 Edition. Ed. Edward N. Zalta. URL = <http://plato.stanford.edu/archives/spr2014/entries/knowledge-acquaindescrip/>

Heisenberg, Werner. 1971. *Physics and Beyond: Encounters and Conversations*. Trans. Arnold J. Pomerans. Harper and Row.

Hillman, Donald J. 1962. "The Measurement of Simplicity". *Philosophy of Science* 29:225-252.

Holmgren, Margaret. 1989. "The Wide and Narrow of Reflective Equilibrium". *Canadian Journal of Philosophy* 19:1:43-60.

Hogarth, William. 1753. *The Analysis of Beauty*. Ed. Joseph Burke. Clarendon Press, 1955.

Hume, David. 1757. *Of the Standard of Taste*. In *The Philosophical Works of David Hume*. Ed. T. H. Green and T. H. Grose. 4 vols. Longman, Green, 1874-1875.

Hutcheson, Francis. 1725. *An Inquiry Concerning Beauty, Order, Harmony, Design*. Ed. Peter Kivy. Martinus Nijhoff, 1973.

Joyce, Richard. "Moral Anti-Realism". *The Stanford Encyclopedia of Philosophy,* Fall 2015 Edition. Ed. Edward N. Zalta. URL = <http://plato.stanford.edu/archives/fall2015/entries/moral-anti-realism/>

Kant, Immanuel. 1788. *Critique of Practical Reason*. Trans. Mary Gregor. Cambridge University Press, 2010.

Kant, Immanuel. 1790. *Critique of Judgment*. Trans. Werner S. Pluhar. Hackett, 1987.

Kant, Immanuel. 1793. *Religion within the Limits of Reason Alone*. Trans. Theodore M. Greene and Hoyt H. Hudson. Harper & Row, 1960.

Kant, Immanuel. 1797. *The Metaphysics of Morals*. Trans. Mary Gregor. Cambridge University Press, 2004.

Kant, Immanuel. 1924. *Lectures on Ethics*. Trans. Louis Inflied. Methuen, 1930.

Kemeny, John. 1953. "The Use of Simplicity in Induction." *Philosophical Review* 62:391-408.

Kivy, Peter. 1976. *The Seventh Sense: A Study of Francis Hutcheson's Aesthetics and Its Influence in Eighteenth-Century Britain*. Burt Franklin.

Kivy, Peter. 1991. "Science and Aesthetic Appreciation". In *Philosophy and the Arts*. Eds. Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein. Midwest Studies in Philosophy, vol. 16. University of Notre Dame Press; pp. 180-195.

Korsgaard, Christine. 1986. "Aristotle and Kant on the Source of Value". In *Creating the Kingdom of Ends*. Cambridge University Press.

Korsgaard, Christine. 1996. *The Sources of Normativity*. Ed. Onora O'Neill. Cambridge University Press.

Korsgaard, Christine. 2009. *Self-Constitution*. Oxford University Press.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. University of Chicago Press.

Kvanvig, Jon. 2003. *The Value of Knowledge and the Pursuit of Understanding.* Cambridge University Press.

Laden, Anthony. 2013. "Constructivism as Rhetoric". In *A Companion to Rawls*. Eds. J. Mandle and D. A. Reidy. John Wiley & Sons.

Laudan, Larry and Jarrett Leplin. 1991. "Empirical Equivalence and Underdetermination". *Journal of Philosophy* 88:449-72.

Mandle, Jon and David Reidy. 2013. *A Companion to Rawls*. Wiley-Blackwell.

Mandle, Jon and David Reidy. 2015. *The Rawls Lexicon*. Cambridge University Press.

Martin, James E. 1989. "Aesthetic Constraints on Theory Selection: A Critique of Laudan". *British Journal for the Philosophy of Science* 40:357-364.

McAllister, James. 1989. "Truth and Beauty in Scientific Reason". *Synthese* 78:25-51

McAllister, James. 1999. *Beauty and Revolution in Science*. Cornell University Press.

McDowell, John. 1983. "Aesthetic Value, Objectivity, and the Fabric of the World". In *Pleasure, Preference and Value: Studies in Philosophical Aesthetics*. Ed. Eva Schaper. Cambridge University Press; pp. 1-16.

Mothersill, Mary. 1986. *Beauty Restored*. Oxford University Press.

Mothersill, Mary. 1989. "Aesthetic Laws, Principles and Properties: A Response to Eddy Zemach". *The Journal of Aesthetics and Art Criticism* 47:1:77-82.

Nozick, Robert. 1981. *Philosophical Explanations*. Harvard University Press.

O'Hear, Anthony. 2000. *Philosophy, the Good, the True, and the Beautiful*. Cambridge University Press.

Osborne, Harold. 1964 "Notes on the Aesthetics of Chess and the Concept of Intellectual Beauty". *British Journal of Aesthetics* 4:160-163.

Penrose, Roger. 1974. "The Role of Aesthetics in Pure and Applied Mathematical Research." *Bulletin of the Institute of Mathematics and Its Applications* 10:266-271.

Plato. *Meno*. Trans. G.M.A. Grube. Hackett, 1980.

Plato. *Republic*. Trans. G.M.A. Grube and C.D.C Reeve. Hackett, 1992.

Plato. *Symposium*. Trans. Alexander Nehamas and Paul Woodruff. Hackett, 1989.

Plotinus. *Ennead*. In *Neoplatonic Philosophy: Introductory Readings.* Trans. John Dillon and Lloyd P. Gerson. Hackett, 2004.

Pritchard, Duncan and John Turri. "The Value of Knowledge". *The Stanford Encyclopedia of Philosophy,* Spring 2014 Edition. Ed. Edward N. Zalta. URL = <http://plato.stanford.edu/archives/spr2014/entries/knowledge-value/>

Popper, K. R. 1959. *The Logic of Scientific Discovery*. Hutchinson.

Quine, W.V. 1960. *Word and Object*. MIT Press.

Quine, W.V. 1969. "Ontological Relativity". In *Ontological Relativity and Other Essays*. Columbia University Press.

Quine. W.V. 1975. "On Empirically Equivalent Systems of the World". *Erkenntnis* 9: 313-328.

Ramsey, Frank. 1931. "Theories". In *The Foundations of Mathematics, and other Logical Essays*. Ed. R. B. Braithwaite. Routledge.

Rawls, John. 1971. *A Theory of Justice*. Harvard University Press.

Rawls, John. 1975. "The Independence of Moral Theory". *Proceedings and Addresses of the American Philosophical Association* 48.

Rawls, John. 1980. "Kantian Constructivism in Moral Theory". *The Journal of Philosophy* 77:9:515-572.

Rawls, John. 1989. "Themes in Kant's Moral Philosophy". In *Kant's Transcendental Deductions*. Ed. E. Forster. Stanford University Press.

Rawls, John. 1993. *Political Liberalism*. Columbia University Press.

Rawls, John. 2001. *Justice as Fairness: A Restatement*. Harvard University Press.

Ross, W.D. 1930. *The Right and the Good*. Oxford University Press.

Santayana, George. 1896. *The Sense of Beauty.* Scribner's.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Harvard University Press.

Scanlon, T. M. 2014. *Being Realistic About Reasons*. Oxford University Press.

Schiller, Friedrich. 1985. *On the Aesthetic Education of Man: In a Series of Letters*. Eds. Elizabeth M. Wilkinson and L. A. Willoughby. Clarendon Press.

Sehon, Scott. 2005. *Teleological Realism*. Bradford Books.

Sencerz, Stefan. 1986. "Moral Intuitions and Justification in Ethics". *Philosophical Studies* 50:1:77-95.

Shaftesbury, Anthony Ashley Cooper, Third Earl of. 1711. *Characteristics of Men, Manners, Opinions, Times etc.* Ed. John M. Robertson. 2 vols. Grant Richards, 1900.

Sibley, Frank. 1959. "Aesthetic Concepts". *Philosophical Review* 68:421-450.

Sidgwick, Henry. 1907. *The Methods of Ethics*. Macmillan.

Singer, Peter. 2005. "Ethics and Intuitions." *The Journal of Ethics* 9:331-52.

Sircello, Guy. 1975. *A New Theory of Beauty*. Princeton University Press.

Sklar, L. 1982. "Saving the Noumena." *Philosophical Topics* 13:49-72.

Sober, Elliott. 1975. *Simplicity*. Clarendon Press.

Sosa, Ernest. 1980. "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge." *Midwest Studies in Philosophy* 5:1:3-26.

Stanford, Kyle. "Underdetermination of Scientific Theory". *The Stanford Encyclopedia of Philosophy*, Winter 2013 Edition. Ed. Edward N. Zalta. URL = <http://plato.stanford.edu/archives/win2013/entries/scientific-underdetermination/>

Stolnitz. Jerome. 1960. *Aesthetics and Philosophy of Art Criticism: A Critical Introduction.* Houghton Mifflin.

Swanton, Christine. 1992. *Freedom: A Coherence Theory*. Hackett.

Swinburne, Richard. 1997. *Simplicity as Evidence of Truth*. Marquette University Press.

Tolstoy, Leo. 1899. "What is Art?" In *What is Art and Essays on Art*. Ed. Louise Maude and Aylmer Maude. Funk and Wagnalls, 1904.

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

Walsh, Dorothy. 1974. "Aesthetic Objects and Works of Art". *The Journal of Aesthetics and Art Criticism* 33:1:7-12.

Walsh, Dorothy. 1979. "Occam's Razor: A Principle of Intellectual Elegance". *American Philosophical Quarterly* 16:241-244.

Weyl, Hermann. 1952. *Symmetry*. Princeton University Press.

White, Morton. 1956. *Toward Reunion in Philosophy*. Harvard University Press.

Wollheim, Richard. 1968. *Art and its Objects.* Harper & Row.

Woodward, James. "Scientific Explanation". *The Stanford Encyclopedia of Philosophy,* Winter 2014 Edition. Ed. Edward N. Zalta. URL = <http://plato.stanford.edu/archives/win2014/entries/scientific-explanation/>

Zagzebski, Linda. 2009. *On Epistemology*. Wadsworth.

Zemach, Eddy M. 1986. "Truth and Beauty". *Philosophical Forum* 18:21-39.

Zemach, Eddy M. 1987. "Aesthetic Properties, Aesthetic Laws, and Aesthetic Principles". *The Journal of Aesthetics and Art Criticism* 46:1:67-73.

VITA

Devon Craig Bryson, who also goes by Devon Craig Brickhouse-Bryson, was born in California in 1987. He completed this dissertation in 2017. He graduated from the University of Tennessee in 2017. He is married to Mary Helen Brickhouse-Bryson.