Abstract: The method of reflective equilibrium starts with a set of initial judgments about some subject matter and refines that set to arrive at an improved philosophical worldview. However, the method faces two, trenchant objections. The Garbage-In, Garbage-Out Objection argues that reflective equilibrium fails because it has no principled reason to rely on some inputs to the method rather than others and putting garbage-in assures you of getting garbage-out. The Circularity Objection argues that reflective equilibrium fails because it has no principled, non-circular way of sorting whatever is put into the method. The moves required to avoid both objections are instructive. Reflective equilibrium requires a meta-justification, and we offer one that appeals to the epistemic goods that underwrite a view known as phenomenal conservatism. Reflective equilibrium calls on us to start with what seems most likely to be true and to alter that collection of judgments in the ways that seem most likely to get us to the truth. Proceeding in this way is epistemically defensible and unavoidable. Hence, reflective equilibrium is not just good, it’s phenomenal.
A PHENOMENAL DEFENSE OF REFLECTIVE EQUILIBRIUM

How do we come to know or justifiably believe philosophical claims? There are a number of competing methodologies in contemporary philosophy. The option closest to counting as the status quo typically goes by the name ‘reflective equilibrium’. For example, this is the method illustrated most often in introductory philosophy courses. It’s also probably closest to the way that philosophers actually reason, despite what they say on paper about being a priori intuitionists, perceptualists, foundationalists, etc.

Despite its widespread deployment, reflective equilibrium has its share of critics. In this paper, we offer a defense of this methodology. First, we show how reflective equilibrium can avoid the two most serious objections in the current literature: the Garbage-In, Garbage-Out Objection and the Circularity Objection. The first argues that reflective equilibrium fails because it has no principled reason to rely on some inputs to the method rather than others and yet putting garbage-in assures you of getting garbage out. The second argues that reflective equilibrium fails because it has no principled, non-circular way of sorting whatever is put into the method.

Second, we argue that the moves required to avoid both objections are instructive. The objections indicate the need for a meta-justification: a justification not of a belief but of a method. Why should we accept that the method of reflective equilibrium improves one’s philosophical worldview? We provide just such a meta-justification that appeals to the same epistemic goods cited in a family of internalist theories of justification known as phenomenal conservatism. Reflective equilibrium calls on us to start with what seems most likely to be true and to alter that collection of judgments in the ways that seem most likely to get us to the truth. Proceeding in this way is both epistemically defensible and unavoidable. Hence, reflective equilibrium is not just good, it’s phenomenal.
I. THE METHOD OF REFLECTIVE EQUILIBRIUM

There are lots of competing philosophical worldviews. We want the one that most closely approximates reality. How do we achieve that? Since at least Goodman, Sellars, and Rawls, one standard answer has been reflective equilibrium. Goodman (1955) applies the method to matters of logic and the determination of the correct rules of inference. He famously wrote that “A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend” (Goodman 1955: 64). In this way, our intuitions about logical rules are balanced against intuitions about the results of particular logical inferences to develop a system of logic that we think it most likely to be correct. Sellars (1962) says much the same thing in philosophy of science. He argues that we ought to manipulate the three components of a world picture—*observed* objects and events, *unobserved* objects and events, and nomological connections—so that there is a maximum coherence among them. And, Sellars notes, “in this reshuffle, no item is sacred” (Sellars 1962: 356).

And finally, Rawls (1999), the philosopher who coined the term ‘reflective equilibrium’, applies the method to matters of ethical and political norms. In his attempt to determine which kind of hypothetical initial position will produce the correct rules of justice, Rawls offers the canonical statement of the method:

[I]f... [our] principles match our considered convictions of justice, then so far well and good. But presumably there will be discrepancies. In that case we have a choice. We can either modify [the initial position] or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth...I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our
considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium. (Rawls 1999: 18)

In all three cases—logic, philosophy of science, and political theory—the method is the same: arrange your philosophical commitments into a coherent framework. For the purposes of this paper, and ease of presentation, we will focus on applications of reflective equilibrium to ethics. In that realm, reflective equilibrium requires that the entirety of one’s moral beliefs (including particular judgments, general rules, and moral schemas) “hang together” in a coherent network.

So far, so good. But how do we go about building such a network? Reflective equilibrium is a method of revision, and such methods need something to start with—something to revise. What data shall we use? In Rawlsian terminology, the input consists of all and only one’s considered moral judgments. These judgments come in the form of particulars (e.g., it was wrong for me to punch Weston) or in the form of more general principles (e.g., it is always wrong to cause suffering). And not just any moral judgment will do. Rawls sets out several conditions for a judgment to count as a considered judgment. For example, judgments must be made in optimal conditions (e.g., the agent must be well-fed, well-rested, not in a heightened emotional state, etc.), the judgment must not be made in self-interest, and the person making the judgment must genuinely desire to have a true belief (Rawls 1999). These conditions ensure that the starting point for reflective equilibrium is as strong as it can be by filtering out epistemically suspect moral beliefs.

Once the data to be sorted has been identified, we need a procedure for sorting these considered judgments. This is where the process of finding an equilibrium begins. There are many theoretical virtues that can be used to sort a pile of judgments. Consistency, yes, but also features like explanatory power and simplicity will likely matter. Sometimes the criteria are described in terms of overall coherency.
Suppose, for example, an inconsistency within a set of moral judgments is detected. If so, the set must be revised in one way or another so that the inconsistency is eliminated. Sometimes this means the rejection of a moral principle. Other times it means the rejection of a judgment about a particular moral case. No individual considered judgment is sacred (Rawls 1999). In the end, the goal is a kind of “smoothing out” of our internal philosophical landscape so that all of our commitments are internally coherent.

What could make the resulting set of beliefs epistemically even stronger? Well, the same thing that can make any intellectual point-of-view stronger: careful reflection on one’s background assumptions, competing alternative viewpoints, and the evidence of others. And this is just what is called for in reflective equilibrium 2.0, sometimes dubbed ‘wide reflective equilibrium’ (Rawls 1974). Again, as applied to morality, a subject examines a wide selection of moral theories as well as all of the relevant arguments for and against them. As with science or politics or religion, thinking carefully about views taken seriously by others can often reveal our own biases, limitations, and oversights. From here on, we’ll let ‘reflective equilibrium’ denote this stronger, wide stance of reflective equilibrium.

The end result is a set of philosophical beliefs made consistent with one another by elimination of the weakest members and sharpened by comparison to the belief sets of thoughtful interlocutors. The process used to compile this set is also repeating and self-correcting. It’s like anti-virus software that runs every night. Reflective people will continue to think about their philosophical commitments and alter them, when appropriate, according to the strictures of reflective equilibrium. What more could you want?

II. TWO OBJECTIONS TO REFLECTIVE EQUILIBRIUM
“Plenty more,” say critics. In broad strokes, what’s needed is a meta-justification of the method of reflective equilibrium itself. In other words, we want to know why reasoning in this way is epistemically appropriate. For example, does engaging in reflective equilibrium increase the odds that we’ll get to a philosophical worldview that is true as compared with a worldview not in equilibrium? If that’s the goal, why think that re-arranging the judgments that one starts with makes it more likely that the resulting set is accurate? To put it baldly, why think that being in reflective equilibrium is an indicator of anything worthwhile?

It’s worth noting that at least some defenders of reflective equilibrium think either that these questions cannot be answered or that reflective equilibrium, properly understood, needs no meta-justification (DePaul 1998; Walden 2013). This is a mistake. Critics have two broad lines of reasoning that independently show reflective equilibrium to be defective. Given these trenchant objections, denying the need for a meta-justification is the philosophical equivalent of putting one’s head in the sand. Something ought to be said in defense of reflective equilibrium. And since virtually every criticism of reflective equilibrium is a species of one of two more basic genera, answering both should put critics back on their heels. The two critiques either object to the input or object to the method of sorting. Let’s have a look at each.

According to what we’ll call the Garbage-In, Garbage-Out Objection, reflective equilibrium has no principled or reasonable way of selecting the inputs to the system (Brandt 1979; Haslett 1987; Williamson 2007; Kelly and McGrath 2010). To see the force of this objection, consider the following thought experiment. Suppose you have a computer program designed to translate English claims into logical form and then test them against one another to build an internally coherent set. Now suppose you want to know something about religion, politics, earth science, or literature. The computer program will help you sort claims against one another. But which claims should you enter into the computer
database to start with? It seems clear that if you dump a bunch of garbage into the machine, you’re bound to get garbage out, albeit garbage that’s arranged in a coherent way.

Our critics say that reflective equilibrium faces a similar quandary. We can’t just dump a bunch of junk into the method and then proceed to rearrange it. If a resolute racist dumps all of his strongly-held racist views into the hopper, you probably won’t like the result. But, on the other hand, if we’re going to restrict the class of opening judgments, we would need a principled way to do so. And saying that only some judgments are eligible for input into the method raises the question of why those judgments are epistemically superior to the excluded others. Any answer to this question would demonstrate that the real epistemic heavy-lifting is done prior to the back-and-forth demanded by method of reflective equilibrium itself.

In fact, the little that Rawls and others say about selecting the initial input appears contrived. Why only considered judgments instead of all judgments? Why exclude judgments that impact our self-interest—do we really think that members of racial minorities should refrain from including judgments about racial equality? And why allow for any judgments for which you have no deeper reason or rationale? In short, the method of reflective equilibrium fails to account for the fact that some starting places are better than others, and without the ability to restrict the starting set of judgments, we run the risk of garbage-in, garbage-out.

Second, according to what we’ll call the Circularity Objection, reflective equilibrium lacks a non-question-begging way of defending its sorting procedure (Haslett 1987; Kappel 2005; Kelly and McGrath 2010). Recall that the method takes initial judgments and brings them into equilibrium by eliminating inconsistencies, cultivating explanatory power, increasing relevance, reducing complexity, etc. But why think that this sort of equilibrium is an epistemic improvement? Only question-begging answers are forthcoming.
The circularity can be seen in two different ways. First, consider again the example of the computer program introduced above. If you are programming a computer to sort an initial pool of data, you’ll have to code the operation to be used in the sorting. But what independent reason can you offer for any given operation or procedure over any of the infinite other sorting procedures available? For example, you might program the computer to aim for theories and principles that are simple rather than complex. Or you might do the reverse. Is there an epistemic reason to prefer one over the other? Reflective equilibrium can only offer a question-begging answer: procedures that inculcate internal coherence are better because they result in belief sets that are more internally coherent.

Here’s a second way to see the circularity of the method. The computer program must sort claims against one another. Take a moral example. If you believe that it’s morally permissible for one boxer to harm another in the boxing ring but also that it’s always wrong to harm another person, you’ve got two moral judgments that are in conflict. How should the conflict be resolved?

If you use the first to disqualify the second, the former claim must have some special epistemic status. Otherwise, it’s just arbitrary that the disqualification goes in one direction rather than another. But in reflective equilibrium, neither beliefs about particulars nor beliefs about principles have priority. Each supports all, and each is supported by all. And so reflective equilibrium is caught in another circle: some initial judgments should be retained at the expense of others but there is no non-question-begging rationale for determining which judgments are primary.

Because of this, say our critics, reflective equilibrium is to be abandoned. Garbage-In, Garbage-Out shows that the method has an arbitrary starting point (or that the epistemic work is being done by some other feature), and Circularity shows that the method has an arbitrary procedure of sorting. Philosophy is better off without reflective equilibrium.
III. PHENOMENALLY DISPATCHING THE OBJECTIONS

The objections to reflective equilibrium are instructive but avoidable. However, the objectors are right about at least this much: avoiding the objections requires telling a deeper epistemic story. The up-side is that this deeper epistemic story provides a meta-justification for reflective equilibrium. The seeds to both the defensive and offensive moves can be found in a passage from Aristotle:

For Plato, too, was right in raising this question and asking ‘Are we on the way from or to first principles?’ There is a difference, as there is in a race-course between the courses from the judges to the turning point and the way back...Presumably, then, we must begin with things evident to us. (Nicomachean Ethics I.4.1095b)

Aristotle’s point is that we can either use our first or basic principles to deduce things about the world or we can take our knowledge of the world to induce our governing first or basic principles. If that pair of options sounds familiar, it should: it’s Roderick Chisholm’s problem of the criterion. We can reason from methods to particulars or from particulars to methods. But which comes first?

Aristotle’s answer is that we should start with what is most evident to us and reason from there. Sometimes what is most evident is a first principle. Other times what is most evident is some particular. But in either case, we start with what we are most sure of. As Aristotle puts it, “we must begin with things evident to us” (Nicomachean Ethics I.4.1095b).

This insight is later developed into a sort of “evident-first” epistemology by modern thinkers like Descartes and Reid, and it finds its clearest contemporary expressions in a family of views known as phenomenal conservatism. These views are ‘phenomenal’ in the sense that the phenomenon of seeming true does all of the epistemic work and ‘conservative’ in the sense that they initially sanction the beliefs one actually holds. In a nutshell, phenomenal conservatism says that it’s reasonable to trust
the way things seem to you unless you have a reason to distrust that seeming. Or, in more analytic form, “If it seems to S as if P, then S thereby has at least \textit{prima facie} justification for believing that P” (Huemer 2001: 99). Catherine Elgin (1996) puts the point in terms of initial epistemic tenability. The beliefs, intuitions, and thoughts that we endorse at any given time are epistemically “innocent until proven guilty.” As she puts it, “being our current best estimates of how things stand, such sentences have some claim on our epistemic allegiance” (Elgin 1996: 101). That claim she calls ‘initial tenability.’

Here’s what this sort of epistemic conservatism looks like in practice. You ask me where my car is parked. I introspect and find that I have a memory of recently parking it in the garage (and no memory of lending it out) so it seems to me that my car is in the garage. If I can find no reason to distrust this seeming, I now have initial justification for the belief that the car is in the garage. While no guarantee, it’s surely the reasonable thing to believe. My belief that my car is in the garage is initially tenable or \textit{prima facie} justified.

Now consider a phenomenal conservative meta-justification of reflective equilibrium. While no philosopher has defended reflective equilibrium along these lines, Scanlon’s recent work on realism about reasons suggests such a defense. Exasperated by critics who press the Garbage-In, Garbage-Out objection, Scanlon concedes that the crucial move in the work of Goodman, Rawls and others is to claim that:

we have reason to believe the things that \textit{seem} to us to be true about these subjects when we are thinking about them under the right conditions, and hence it is appropriate for us to treat these beliefs as considered judgments in a process of seeking reflective equilibrium if we have no apparent reason to doubt them. (Scanlon 2014: 82-83, emphasis ours)
Phenomenal conservatism explains why this is the right thing to say: the fact that P seems true is a reason to think that P is true in the absence of defeaters. So then how should we respond to each of the major objections to reflective equilibrium?

Let's begin with the Garbage-In, Garbage-Out objection. The question is where to start. Out of all possible propositions, which ones should we put into the sorting machine? The reply from phenomenal conservatism is straightforward: start with all and only the claims that seem true to you. After all, we have to start somewhere, and there is no better place to begin. For example, as applied to a moral case of reflective equilibrium, we should start with all and only the moral judgments that seem true. This starting point is non-arbitrary since phenomenal conservatism sanctions seeming true as an epistemically relevant property—the very thing demanded by the Garbage-In, Garbage-Out Objection.

Next consider the Circularity Objection. Recall that the problem is that reflective equilibrium offers no way of determining which sorting procedures are better than others and which beliefs should give way to others in the case of an inconsistency. Phenomenal conservatism offers an antidote to both problems. First, some sorting procedures seem more likely to result in truth than others. Those are the procedures we should use to sort the initial data. In other words, the same way that we have seemings about the fodder, we also have seemings about the procedures. In Chisholm’s terminology, we have seemings about both particulars and methods. For example, we have a strong intuition that contradictions can't be true. This underwrites a procedure requiring consistency. And so phenomenal conservatism has a non-question-begging explanation for why some sorting procedures are to be preferred to others.

Second, phenomenal conservatism helps us in cases where two initial judgments come into conflict. The relative strength of a seeming will dictate which judgment is the defeater and which judgment is defeated. In terms of the computer program example introduced earlier, suppose every
claim loaded into the initial database comes with some meta-datum: a number between 0 and 1 that represents the “strength” of the judgment. This gives the computer an independent way to specify the outcome of conflicts between judgments in the initial set. Suppose, for example, the claim \( \sim P \) has a meta-data number of .6 and P has a meta-data number of .8. Rejecting either P or \( \sim P \) will bring the set into consistency, but the strength of judgments means that—other things being equal—the more coherent set will be the one that rejects \( \sim P \). This means that sometimes even judgments that we are initially very confident in will later be revised to produce a more coherent, overall set of beliefs.1

The translation to philosophical inquiry is straightforward. During a sorting procedure, what gets tossed depends on the strength of one’s philosophical convictions, and the difference in credence gives us our starting rank of judgments. Some things seem true whereas others seem false. Among the things that seem true, some seem more likely to be true than others. That seeming state is our epistemic touchstone. As a moral example, if you are sure that it is always wrong to harm a child, then you’ll have to give up the particular judgment that it was morally permissible to use atomic weapons at Hiroshima. This offers a non-arbitrary way to rate a judgment’s resistance to revision.

So phenomenal conservatism can solve both sides of the circularity objection. And, as the cat in the hat says, that’s not all. Phenomenal conservatism nicely captures the thought that both our initial judgments and our initial procedures are subject to revision. In other words, seeming states can be invoked in a kind of meta-evaluation of the standards we are using at any given time. For example, as applied to morality, many philosophers have argued that our moral intuitions are defeasible and that our initial starting points can and should be impacted by evidence of bias and evidence of successful social arrangements (e.g., Anderson 2015).

This is not just a hypothetical problem. Over the last fifteen years, experimental philosophers have tallied a long list of ways that initial philosophical judgments go awry. There is good evidence that
philosophical intuitions and initial philosophical judgments are sensitive to all sorts of irrelevant factors. Sometimes these factors are cultural. For example, Weinberg et al. (2001) show that judgments about Gettier cases differ with a Western or Asian upbringing, and Machery et al. (2004) proves the same about intuitions concerning reference and the semantic content of names. Sometimes these factors are gender or race specific. Zamzow and Nichols (2009) demonstrate that men and women have different moral intuitions about trolley cases. Uhlman et al. (2009) provide evidence that race affects moral judgments in lifeboat cases. Finally, our philosophical intuitions are also subject to order and framing effects, both in epistemology (Swain et al. 2008) and ethics (Petrinovich and O’Neill 1996).

While troubling, these findings don’t sink the reflective equilibrium plus phenomenal conservative program. That’s because reflective equilibrium can accommodate higher-order considerations (e.g., Walden 2013, Greenspan 2015), and phenomenal conservatism can explain why it’s reasonable to trust the evidence showing that some intuitions are faulty. If it ever seems to you that some of your judgments are biased or that some of your sorting criteria are not truth-indicative, this higher-order evidence will affect the input to the method, the sorting procedures of the method, or both.

In short, phenomenal conservatism provides us with a non-arbitrary starting point (since some propositions will have initial justification), a non-arbitrary procedure for sorting the initial judgments (since some methods of sorting will seem more likely to get to the truth than others), and a non-arbitrary way of refining the method as time goes along. The objectors have the non-question-begging answers that they demanded.

No doubt this defense of reflective equilibrium will not satisfy everyone. For example, one might respond in the following way: “Look, it now seems that phenomenal conservatism is doing all of the
epistemic work: we’re justified in believing any claims that seem true to us, full stop. There’s no room for reflective equilibrium to do anything further.”

But that’s not true, or at least it’s not true for most of us. Most of us have inconsistent philosophical worldviews. And most of us think that such inconsistencies are problematic. Most of us think that we are biased in certain ways or vulnerable to irrelevant influences, etc. And, so most of us have internally-accessible reasons to engage in reflective equilibrium: it seems to us that doing so improves the odds that our philosophical worldviews are correct. In other words, phenomenal conservatism explains how believing certain things at any given moment meets our *synchronic* epistemic goals. But reflective equilibrium explains how a process meets our *diachronic* epistemic goals. There is still a place for reflective equilibrium in the overall epistemic story.

Second, one might object to this solution offered here in the following way: the fact that something seems true to you is no guarantee that it is actually true. And, so you might still end up with garbage going into the reflective hopper, albeit garbage that seems okay to you. This objection is fine as far as it goes, but it doesn’t go very far, at least with non-skeptical crowds. That’s because the objection is asking for yet a further meta-justification of phenomenal conservatism: why think that seeming true is an epistemic virtue or gets us what we want? While a full-fledged defense of phenomenal conservatism is beyond the scope of this paper, we grant that a phenomenal conservative defense of reflective equilibrium is only as good as our reasons for endorsing phenomenal conservatism. We close by noting three things.

First, there are a variety of plausible defenses of phenomenal conservatism in the literature (see Tucker 2013 for an overview). Some are practical: because our beliefs work for us, they are likely to be true. For example, Elgin argues “that the sentences we accept do not in general frustrate our efforts is some reason to accept them” (Elgin 1996: 102). Other defenses are philosophical. For example,
Huemer (2007) argues that a belief is justified only when it is based on that which does the justifying and that—as a matter of fact—all human beliefs are based on the way things seem to us. So, if our beliefs are justified at all, they are justified by seeming states. Given the existence of these sorts of defenses, it is reasonable to deploy phenomenal conservatism in defense of reflective equilibrium.

Second, the methodology of reflective equilibrium bolstered with phenomenal conservatism is epistemically on par with other, exemplary models of knowledge. Take the paradigm case of science. As Sellars makes clear, in the domain of science, the seemings provided by our interaction with the world plus our seemings about ways to organize this data, provide us with empirical conclusions about the world. For example, it seems to us that the world must be consistent, and so we reevaluate observations that do not fit with our initial theories or previous observations. Even acceptance of the scientific method relies on a seeming state: it seems like a reasonable way to go about discovering things about the world.

You would be forgiven for thinking that science relies not on seemings but on empirical observations. But that is merely kicking the can down the road: *why* do observations of the world matter? Well, because they provide powerful seeming states that we then systematize through the scientific method. The actual contact with the world does no epistemic work: scientists in the Matrix or the evil demon world have justified beliefs, too.

Third, we think it’s likely that something like phenomenal conservatism plus reflective equilibrium is the only game in town. The fact is that all principles and methodologies have to stop somewhere—we can’t have a metajustification for everything. Alston (1996) convincingly shows that defending any doxastic practice, even perception, requires either an infinite number of belief-forming methods (so that method A can be defended by appeal to method B, B can be defended by appeal to method C, and so on) or else epistemic circularity. Neither is a happy ending.
Given our situation, what can we do other than start with what seems to be true about the world and go from there? There are no other options. So recent defenders of reflective equilibrium were half-right: the search for reasons does, indeed, “bottom out.” But it does so in phenomenal conservatism rather than reflective equilibrium. It’s the former that justifies our use of the latter.

In sum, the addition of phenomenal conservatism shows why the method of reflective equilibrium is epistemically defensible. Rawls was right that we ought to do our best to form a constellation of commitments held together in a reflective equilibrium. He just didn’t know why.


1A very helpful referee raises the following question: and what ought to happen when two initial judgments are held equally strongly and yet are inconsistent? The first question to ask about this case is whether each of the initial judgments cohere equally well with the rest of what seems true at the time. If not, the one that coheres less well should be rejected. That’s because phenomenal conservatism justifies our believing that things are the way they seem until we have reason to think otherwise. And so it’s always the case that the more minor belief revision is to be favored over the more major revision, where ‘major’ and ‘minor’ are a function of strength of seemings for our body of belief. If the two, inconsistent beliefs are held equally strongly and cohere equally well with the body of propositions that seem true, then there are no epistemic considerations that would mitigate against accepting one over the other. Accepting one over the other would be purely arbitrary from an epistemic point of view.

2The authors would like to thank Dugald Owen and Sarah Roberts-Cady for many helpful conversations about the moves in this paper and for their careful comments on the initial drafts of the ideas. Some of these ideas were developed in Sarah Roberts-Cady's senior seminar course on the work of John Rawls. Thanks to her students for valuable feedback on proto-versions of the ideas in this paper. Thanks to the two anonymous referees for the Journal of Philosophical Research for excellent suggestions for improvement. Some of this research was completed while one of the authors, Justin McBrayer, was on a Fulbright appointment to the University of Innsbruck, Austria. Many thanks to the taxpayers who support the Austrian-American Fulbright Commission and the Fulbright staff who made this research possible.