The Demands of Epistemic Rationality: Permissivism and Supererogation

By Han Li
B.A., Cornell University, 2011

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in the
Department of Philosophy at Brown University

Providence, Rhode Island
May, 2018
© Copyright 2018, by Han Li
This dissertation by Han Li is accepted in its present form by the Department of Philosophy as satisfying the dissertation requirement for the degree of Doctor of Philosophy

Date: ________________

David Christensen, Advisor

Recommended to the Graduate Council

Date: ________________

Joshua Schechter, Reader

Date: ________________

Christopher Meacham, Reader

Approved by the Graduate Council

Date: ________________

Andrew Campbell, Dean of the Graduate School
VITA

Han Li studied both philosophy and government at Cornell University, receiving his B.A. in 2011. Since starting his Ph.D. in the Department of Philosophy at Brown University, he has worked mainly in epistemology, having published work in the journals Synthese, Episteme, and Erkentniss. He has also taught classes in epistemology, formal logic, decision theory, the philosophy of probability, ethics, and introduction to philosophy. He is currently Visiting Assistant Professor in the Department of Philosophy at Kansas State University.
ACKNOWLEDGEMENTS

I would like to thank all the great teachers who have helped me complete this dissertation, including Jamie Dreier, Nina Emery, Christopher Meacham, and Joshua Schechter. I would especially like to thank my advisor David Christensen. I will be forever grateful for the wisdom, encouragement, and patience he has shown me throughout my many years at Brown. David is truly the best teacher I have ever had.

I would also like to thank my many friends and colleagues, both past and present, who have made my time at Brown not just intellectually satisfying, but also a whole lot of fun.

Finally, I would like to thank my family, especially my parents (who never complained too loudly about the whole “philosophy” thing) and Kelsey Brykman (who has put up with living with a philosopher for already too many years).
TABLE OF CONTENTS

THE TROUBLE WITH HAVING STANDARDS ................................................................. 1
1. Introduction ............................................................................................................ 2
2. The ESV – Preliminaries ...................................................................................... 3
3. Two Motivations .................................................................................................. 6
   4.1. The Normative Criterion ............................................................................... 10
   4.2. The Applicability Criterion .......................................................................... 12
5. The Belief Theory of Standard Possession ........................................................ 14
6. The Dispositional Theory of Standard Possession .............................................. 16
   6.1. Dispositions and Performance Errors ......................................................... 16
   6.2. Idealized Dispositions .................................................................................. 18
   6.3. Dispositions and the Normative Criterion .................................................... 20
7. No Explanation Needed ....................................................................................... 23
8. Conclusion – A Diagnosis ................................................................................... 27
9. Works Cited ......................................................................................................... 29

HOW SUPEREROGATION CAN SAVE INTRAPERSONAL PERMISSIVISM ............ 32
1. Introduction ............................................................................................................ 32
2. Intuitive Considerations ....................................................................................... 38
3. The Arbitrariness Objection ................................................................................. 39
4. There Are No Cases of Known Permissivism ...................................................... 41
   4.1. The View ....................................................................................................... 41
   4.2. The Problem .................................................................................................. 45
5. Epistemic Standards and Accuracy .................................................................... 46
   5.1. The View ....................................................................................................... 46
   5.2. The Problem .................................................................................................. 46
6. A Diagnosis .......................................................................................................... 50
7. Embracing Supererogation .................................................................................. 52
8. Conclusion .......................................................................................................... 54
9. Works Cited ......................................................................................................... 54

A THEORY OF EPISTEMIC SUPEREROGATION ......................................................... 56
1. Introduction ............................................................................................................ 56
2. The Good and the Right ...................................................................................... 60
3. Requirements and Epistemic Virtues .................................................................. 61
4. “Coming Up” with Hypotheses .......................................................................... 66
5. Requirements and Creativity ............................................................................. 69
6. Too Much Supererogation? ................................................................................ 72
7. Epistemic Justification and Good Housekeeping .............................................. 76
8. Alternative Theories ............................................................................................ 79
9. Conclusion .......................................................................................................... 83
10. Works Cited ....................................................................................................... 83
THE TROUBLE WITH HAVING STANDARDS

Abstract: The uniqueness thesis states that for any body of evidence and any proposition, there is at most one rational doxastic attitude that an epistemic agent can take toward that proposition. Permissivism is the denial of uniqueness. Perhaps the most popular form of permissivism is what I call the Epistemic Standard View (ESV), since it relies on the concept of epistemic standards. Roughly speaking, epistemic standards encode particular ways of responding to any possible body of evidence. Since different epistemic standards may rationalize different doxastic states on the same body of evidence, this view gives us a form of permissivism if different agents can have different epistemic standards.

Defenders of the ESV, however, have not paid sufficient attention to what it means to have a particular epistemic standard. I argue that any theory of epistemic standard-possession must satisfy two criteria to adequately address the broader needs of the ESV. The first criterion is the normative criterion: a theory of standard-possession should explain why agents are rationally required to form beliefs in accordance with their own (rational) epistemic standard, rather than any other (rational) standard. The second criterion is the applicability criterion: a theory of standard-possession should rule that agents have the epistemic standards we intuitively think they have. I then argue that no extant theories of standard-possession can satisfy both these criteria. I conclude by diagnosing why these criteria are so hard to jointly satisfy. Defenders of the ESV are thus left with a serious obstacle to forming a complete and plausible version of their view.
1. Introduction

The uniqueness thesis states that for any body of evidence and any proposition, there is at most one rational doxastic attitude that an epistemic agent can take toward that proposition. Permissivism is the denial of uniqueness. A certain type of permissive theory stands as perhaps the most popular. This is the theory that I will call the Epistemic Standard View (ESV), since it relies on the concept of epistemic standards. Roughly speaking, epistemic standards encode particular ways of responding to any possible body of evidence. If more than one epistemic standard is rational to have, then different agents can have different epistemic standards while remaining rational. So for those bodies of evidence where different rational standards will advise different doxastic responses, there is more than one doxastic response to that evidence that is consistent with rationality. Thus, it seems that we get a plausible form of permissivism.

As we will see, central to this way of thinking about permissivism is the claim that epistemic agents have some particular epistemic standard. If the agent’s epistemic standard is rational, then the agent is not only permitted to form beliefs in accordance with it, but she is in fact required to do so. This is true even if there are other standards which are also rational. Thus, if the concept of epistemic standards is to deliver a plausible version of permissivism, we need to understand what it means to have a specific epistemic standard. That is, we need a theory of epistemic standard-possession.

---

1 See (White 2013) and (Schoenfield 2013) for representative formulations of the uniqueness thesis. “Doxastic state” here can refer to either all-or-nothing beliefs or partial beliefs – these will result in different versions of the thesis. There is no version of the uniqueness thesis where “doxastic state” can refer to both all-or-nothing beliefs and partial beliefs (uniqueness is not supposed to rule out agents both rationally believing P and rationally having a high credence in P, for example).

2 Paradigmatic examples of the ESV in general include (Schoenfield 2013) and (Titelbaum and Kopec ms). More specific versions include Subjective Bayesianism and the view attributed to William James in (Kelly 2013). As we will see later, many permissive writers employ aspects of the ESV when trying to elucidate or defend particular features of permissivism.
On the final analysis, however, formulating a theory of epistemic standard-possession that can comfortably support the type of permissive theory we want turns out to be extremely difficult. To see this, we will begin in §2 by sketching out the ESV. Then, in §3 we will examine two motivations for the ESV in depth. We will see how the concept of epistemic standard-possession is employed to satisfy these motivations. Next, in §4 we will then look at two criteria for a theory of standard-possession. These are the criteria that a theory must satisfy if epistemic standards are to be employed in a way that answers to the original motivations. In the next two sections (§5 and §6), we will go over two broad families of theories of standard-possession, and find each of them unable to satisfy both criteria. After the failures of these theories, we might conclude that there is something wrong with our criteria for an adequate theory. So in §7 we will examine a skeptical response to the challenge laid out in this paper, and find the solution wanting. Finally, we conclude in §8 by trying to diagnose why all our attempts at a theory of standard-possession were ultimately failures. This will supply us with an in-principle reason for thinking that it will be difficult to develop a theory of standard-possession that can satisfy the needs of the ESV. At the very least, defenders of the ESV are left with an onerous challenge.

2. The ESV - Preliminaries

We will begin by sketching out the essential aspects of the ESV, as it will be understood in this paper. First, we will think of epistemic standards as a certain type of abstract object – functions from total bodies of evidence to doxastic states. When inputted with a total body of evidence, epistemic standards output a complete doxastic state.\(^3\) This is just for ease of

---

\(^3\) In thinking of epistemic standards as functions, I am mainly following (Schoenfield 2013, p. 7).
exposition, as nothing in this paper will crucially depend on the metaphysics of epistemic standards. Some of these functions have a special relationship to particular epistemic agents. This is because every (or at least almost every) epistemic agent possesses a particular epistemic standard.

By thinking of epistemic standards in this way, we can separate the question of what an epistemic standard is from the question of what it takes to have a particular epistemic standard. Many proponents of the ESV do not make this distinction, since their conception of an epistemic standard comes with a conception of what it takes to have an epistemic standard “built in”. 4 Since we will be canvassing many different versions of the ESV, making this distinction will be helpful for zeroing in on the exact question we are interested in.

According to the ESV, epistemic standards themselves can be assessed for rationality. For the theory to be permissive there must be multiple distinct standards that are rational to have. Most defenders of the ESV, however, still claim that there are some rational requirements on the epistemic standards themselves. 5 Some writers – subjective Bayesians being a notable example – think of these as nothing more than internal consistency requirements. Others go further and think that there are also substantive requirements on epistemic standards (requirements which rule out counter-induction, for example). In either case, what is important for our purposes is that forming beliefs in accordance with irrational standards always results in irrational beliefs.

Of the rational requirements on epistemic standards, there is one worth noting for our purposes: immodesty. An epistemic standard is immodest just in case it recommends itself as the

---

4 If we think of epistemic standards as a specific type of belief, for example, then it is obvious how agents have epistemic standards – after all, beliefs only exist if they are possessed by agents.
5 A possible exception is the theory presented in (Foley 1987).
most truth-conducive standard available to the agent.\textsuperscript{6} The idea is that epistemic agents are required to believe that their own standard is the most reliable standard that they can use – the one most likely to get to true beliefs. There seems to be something \textit{incoherent} about having a particular standard while thinking that some other standard is more likely to output true beliefs. So rational epistemic agents with rational standards will always believe their own standard to be uniquely truth-conducive. Immodesty is perhaps best thought of as an internal consistency requirement on epistemic standards, and almost every version of the ESV has such a requirement in place.\textsuperscript{7}

If an agent has a rational epistemic standard, which beliefs are rational for her depends crucially on which epistemic standard she has.\textsuperscript{8} The ESV is a theory which conceives of the evidential support relation as a three-place relation instead of a two-place relation. Evidence does not support some belief full stop, but only relative to the agent’s (rational) epistemic standard. These standards vary from agent to agent, and so the same body of evidence might support different beliefs for different agents. For a single agent with some specific standard, however, the standard only relates a body of evidence to a unique doxastic state. So the agent is epistemically required to adopt the doxastic state which her epistemic standard outputs given her total evidence. As a result, there is only one belief that any one agent can be fully rational in

\textsuperscript{6}More precisely, an agent’s epistemic standard is immodest just in case it outputs beliefs that maximize \textit{expected accuracy} from the agent’s own point of view, compared to the beliefs outputted by any rival standard the agent can adopt. Expected accuracy is a measure of how close an agent can expect some set of beliefs to be to the truth. Though this notion is usually defined in terms of credences, it is natural enough to discuss an analogous notion for all-or-nothing beliefs. See (Lewis 1971) and (Moss 2011) for examples of how to understand immodesty in terms of credences and the notion of expected accuracy.

\textsuperscript{7}See (Horowitz 2014) for a further discussion on immodesty as it relates to epistemic standards.

\textsuperscript{8}It is an open question what agents with irrational epistemic standards should do. One possibility is that these agents are constantly in epistemic dilemmas, with nothing they do being epistemically rational. Another possibility is that these agents are required to somehow change their epistemic standard to a rational one. This response seems to run into problems with the arbitrariness objection, described below. In any case, this difficult question is irrelevant for our purposes.
holding toward a single proposition on any body of evidence.\(^9\)

### 3. Two Motivations

With these details in place, we are now equipped to examine two motivations for the ESV. This will not only allow us to see the allure of such a view, it will set us up to see problems with it down the road. The two motivations are not exclusive to the ESV – they are behind most, if not all, permissive theories of rationality. Still, the ESV is tailor-made to fulfil these two motivations, and it does so in a particularly satisfying way.

The first motivation for the ESV involves a general difficulty for permissivism: the arbitrariness objection. This objection is an obstacle that must be overcome for the existence of permissive cases to be plausible at all.\(^10\) The problem is that if you know you are in a case that is permissive between two different beliefs, then choosing one belief over the other seems arbitrary. There is, after all, no reason to pick any one particular belief. But forming a belief for no reason seems irrational. And if forming either belief is irrational, how can this case be permissive in the first place?

To make this point vivid, imagine a rational agent who has two pills – one which induces a belief in P, and one which induces a belief in ~P.\(^11\) Clearly, it would be irrational for her to form a belief about P by taking one of the pills at random. This would be tantamount to forming beliefs by flipping a coin, at least with regards to getting at the truth. And forming beliefs by coin-flip is irrational if anything is. But now suppose the agent’s situation is permissive between a belief in P and a belief in ~P, and the agent knows this. So why should the agent not just pop a

---

\(^9\) See (Schoenfield 2013, pp. 8–9) for an in-depth presentation of this general picture.

\(^10\) Many variations of this problem were first suggested by White in his (2005) and then his (2013). Other versions of this worry are presented by (Christensen 2007) and (Feldman 2007).

\(^11\) The belief-inducing pill example is due to (White 2005, 2013).
pill at random? After all, she is guaranteed to end up with a belief supported by her evidence (and she knows this). So permissivists are left without an explanation of why taking the pill would be irrational. And after she does take a pill, why not take the pill for the contradictory belief? Again, either way she will end up with a belief supported by her evidence. This situation is clearly absurd. Yet if permissivism is true, the absurdity is hard to explain.

The ESV has a satisfying solution to this problem. Let us imagine that the pills not only induce a belief in either P or ~P, but they also induce in the agent whatever epistemic standard would make the induced belief rational given her evidence. A rational agent, says the defender of the ESV, would not be interested in such a pill. This is because from her perspective, the other standard is not as good as her own. Her own standard, after all, is rational. It is therefore immodest. Thus, the agent must believe that her epistemic standard is especially truth-conducive (even though she may recognize other standards as equally rational).

Since any one epistemic standard recommends a unique doxastic state in any situation, an agent’s epistemic standard advises the agent to form one of the two doxastic attitudes toward P – not both. Whatever attitude is recommended, she should regard it differently from the contradictory attitude, even if she knows both beliefs are rationally permitted. This is because the agent should think that the belief recommended by her epistemic standard – what she believes to be the more truth-conducive of the two standards – is the one that is most likely to be true. Thus, forming a belief by considering the evidence and consulting her own standard is better than taking a pill at random. Though both methods will result in a rational belief, one method is more likely to result in a true belief, from the agent’s point of view.12 Thus, the

12 See (Schoenfield 2013, pp. 8–9) for a paradigmatic version of this type of reply. This way of answering the arbitrariness objection is extremely popular in the literature on permissivism. Other examples include (Douven 2009), (Weintraub 2013), (Meacham 2013), (Kelly 2013), and (Podgorski 2016). The ESV is also mentioned by
arbitrariness objection is put to rest.

The second motivation for the ESV is perhaps the single most cited reason for holding any type of permissive view. The following passage from Gideon Rosen (2001) is often quoted approvingly by permissivists:

It should be obvious that reasonable people can disagree, even when confronted with a single body of evidence. When a jury or court is divided in a difficult case, the mere fact of disagreement is does not mean that someone is being unreasonable. Paleontologists disagree about what killed the dinosaurs. And while it is possible that most of the parties to this dispute are irrational, this need not be the case. To the contrary, it would appear to be a fact of epistemic life that a careful review of the evidence does not guarantee consensus even among thoughtful and otherwise rational investigators (7).

In the actual world we see a wide range of opinion on all sorts of difficult topics. This often results in disagreements between intelligent and well-meaning people, even if they have the same body of evidence. Intuitively, it seems that many of these epistemic agents can be holding rational beliefs. But if uniqueness is true, any time two agents disagree with the same body of evidence, at least one of them must be believing irrationally. So to avoid this overly harsh verdict, many writers embrace permissivism.

While this is a widely shared motivation among permissivists, defenders of the ESV have a particular way of explaining how this kind of disagreement arises. Consider the example of the disagreements about the existence of God that arise between atheists and theists. Presumably,
there are atheists and theists in the same evidential state. Perhaps they are both well acquainted with all available evidence – the contents of holy books, facts about religions and their history, archaeological and geological records, philosophical arguments on either side, etc. According to the ESV, the two well informed agents may yet have different epistemic standards. If one standard – the atheist standard – advises disbelieving in the existence of God on their shared evidence, then anybody who has that standard is rational in forming such a belief. Similarly, someone with the theist standard can be rational in believing that God exists. So the two agents can disagree while both being rational.

There is something intuitive about this explanation. When we fully flesh out our imagined atheist and theist, it is also natural to imagine that they are different people with different backgrounds. The theist, for example, might have been raised by theist parents, or attended church regularly as a child, or gone to a private religious school. It is intuitive to think that these types of factors made the theist a very different type of reasoner from someone who was raised as an atheist. And it is this difference – a difference in how the agents reason – that ultimately explains their difference of opinion. The ESV captures this intuition by cashing out different ways of reasoning as epistemic standards. This is another point in favor of the view.


It should be clear that the ESV relies heavily on the concept of having a particular epistemic standard. That is, not only do there exist different epistemic standards, but agents

---

Other plausible alternatives include disagreements about controversial subjects in philosophy or science (such as Rosen’s example of paleontology).

16 One might think that having different backgrounds is incompatible with having the same evidence. We should not, however, think that the sort of subtle effects that past experiences have on our belief-forming behavior constitute evidence. For all sorts of things affect the way actual humans form beliefs. If all of these count as evidence, then two agents never have the same evidence in real life cases. This is intuitively unacceptable. For a more detailed discussion of this point, see (Schoenfield 2013, pp. 4–5).
somehow possess a particular standard, to the exclusion of all others. This is why an agent can non-arbitrarily form a belief about P in a situation, even if she knows the situation to be permissive with regards to P – her own standard will advise a unique attitude toward P. And this is why the atheist and the theist can disagree about the existence of God – they have different epistemic standards advising different beliefs on the subject.

Thus, in order to flesh out the ESV, we need an understanding of what it means for an agent to have some specific epistemic standard. There is presumably some fact about an agent that makes it the case that the agent has the standard she has, rather than some other standard. We need a “theory of standard-possession” that can tell us which facts these are. This is an issue that has received surprisingly little attention in the literature.

When searching for a theory of standard-possession, there are two separate criteria that we should keep in mind. Satisfying these criteria is necessary for the ESV to be responsive to the motivations presented in §3. We will call the first criterion the normative criterion, and the second criterion the applicability criterion.

4.1. The Normative Criterion

The normative criterion is that a theory of standard-possession should explain why an agent is required to form beliefs in accordance with her own rational epistemic standard, rather than some other rational standard. To understand the nature of this criterion, it will be helpful to contrast it with a related but different explanatory demand, viz. an explanation for why an agent should believe in accordance with a rational standard at all. This more general demand is not specific to the ESV, as even defenders of uniqueness need to answer it. Any theory of rationality needs to explain the peculiar normative force of the concept. But even if we understand why we
should be rational rather than not, defenders of the ESV have an additional explanatory burden, viz. why we should be rational in some particular way rather than any other (equally rational) way. The ESV says that there are many different rational epistemic standards, yet every agent is rationally required to believe in accordance with their standard. This is how the ESV answered the arbitrariness objection. So there is clearly some normative force grounded in the possession of a rational standard. An adequate theory of standard-possession should explain this normative force.\textsuperscript{17}

One might think that such an explanation can be found simply in the idea that epistemic standards must be immodest in order to be rational. Recall that immodest standards are those which recommend themselves as the most truth-conducive standard available, and that immodesty is a coherence constraint on epistemic standards. But it is not mysterious why rational standards must be coherent. And it is also not mysterious why agents must believe according to the standard that they think is most truth-conducive. So, the thought goes, there is nothing left to explain.

This line of thought goes too fast. Even if we can understand why rational standards must be immodest, this only shows us why an agent who follows her own standard will believe that her standard is especially truth-conducive. But she is only rationally forced to have this belief if she is rationally required to believe according to her own standard rather than some other standard. This is exactly what we are trying to explain in the first place. So, on pain of circularity, the immodesty of rational standards cannot explain why an agent must form beliefs

\textsuperscript{17} In addition, it is worth noting that this criterion is distinct from the question of why an agent cannot rationally change her standard. Even if agents can change their standards, one might still contend that an agent who has some standard S while forming beliefs which are not in accordance with S is making a rational mistake. This latter claim is what is currently at issue. While the question of changing epistemic standards is an interesting one, I will put the issue aside for the purposes of this paper.
according to her own standard.

To see the problem more clearly, we can think about an obviously absurd theory of standard-possession. For example, imagine a theory under which an agent’s epistemic standard is grounded in the color of her hair. This theory claims that there are two rational standards – the black hair standard and the blonde hair standard. (Agents with any other color hair are irrational). Imagine an agent with black hair who believes that the blonde hair standard is the most truth-conducive standard available. This belief is supposedly irrational. But why is it irrational? The standard explanation is that it is irrational because it does not cohere with having the black hair standard. But, under this theory, having the black hair standard is a matter of having black hair. And clearly no beliefs are incoherent with having black hair – this is just nonsense. Hair color does not seem to be the type of thing that can cohere (or fail to cohere) with beliefs. So it makes no sense to think that agents are required to believe that their own standards are especially truth-conducive under the hair color theory of standard-possession. This is enough to make it a bad theory.

This is not to say that talk of immodesty is necessarily pointless. If a theory of standard-possession does allow us to understand why having a particular standard is incoherent with believing that the standard is not truth-conducive, then we may end up with a good strategy for satisfying the normative criterion. The point is just that we can’t assume such a strategy will work without actually looking at the details of what it means to have a particular standard.

Finally, one might wonder whether this normative criterion is actually a fair constraint on theories of standard-possession. Is this the type of thing that even calls out for explanation? I will consider this objection in §7.
4.2. The Applicability Criterion

The idea behind the applicability criterion is that a theory of standard-possession should apply to the cases, both actual and merely possible, that we intuitively want it to apply to. A good theory of standard-possession should rule that all, or at least the vast majority of, actual (non-ideal) agents have epistemic standards in the first place. And it should rule that agents have the particular standards that we intuitively think they do.

To make this more concrete, recall our case of rational disagreement between the atheist and the theist. The theory we end up with should make it plausible that the atheist has an atheist epistemic standard while the theist has a theist epistemic standard – that was one of the original motivations for the ESV. This means there must be something about atheists that makes it the case that they have different standards than theists have, and this something should play an explanatory role in their actual disagreement. And intuitively, this difference should be grounded in facts about how these agents actually reason. More generally, there must be some facts about how an agent forms her beliefs which at least co-vary with which epistemic standards she has in a comprehensible way. Again, the hair color theory of standard-possession would fail this criterion, since there is no correlation between hair color and belief in God.

Given this requirement, a good theory of standard-possession should also allow for the possibility of agents having irrational standards. After all, we can imagine fictional agents who reason in all sorts of crazy ways. Imagine, for example, an agent who was raised by an affirming-the-consequent cult. She reliably and consistently affirms the consequent whenever she can. She says things like “affirming the consequent is the best way to get true beliefs!” When she meets others who fail to affirm the consequent she attempts to “correct” them. Intuitively, it seems that this agent has an affirming-the-consequent standard of reasoning. Thus,
the same reasons we have for saying that the atheist has “atheist standards” and the theist has
“theist standards” will also force us to say that this agent has “affirming-the-consequent
standards.” Since standards that endorse affirming the consequent are irrational if any standards
are, we see that it must be at least possible for agents to have irrational epistemic standards. A
good theory of standard-possession should be consistent with this possibility.

Finally, it is important to note that ordinary human beings can be irrational in more
mundane ways. That is, they are prone to reason in ways that are counter to their own epistemic
standards. After all, even practicing logicians will occasionally make invalid inferences. A good
theory of standard-possession should imply that such occasions are in fact performance errors,
rather than implying the logician has an epistemic standard which endorses invalid inferences.

5. The Belief Theory of Standard-Possession

With our two criteria in place, we can begin our review of possible theories of standard-
possession. We will begin by considering one of the few views explicitly proposed by defenders
of the ESV. The idea is simply to identify having a particular epistemic standard with the belief
in the truth-conduciveness of that standard.\textsuperscript{18} So an agent’s belief in the truth-conduciveness of
that standard is not just \textit{rationalized} by her having that standard – it is what it means for her to
have the standard in the first place.

This theory seems to deftly satisfy the normative criterion. The standard immodesty
strategy works here – indeed, the argument becomes trivial. It is clear why having a standard S
is incoherent with not believing that S is the most truth-conducive standard available.

\textsuperscript{18} (Schoenfield 2013) and (Elga ms) express versions of this theory. A related theory is a version of subjective
Bayesianism where epistemic standards are just an agent’s initial priors, understood as actual (psychologically real)
credences toward every sentence in the language. This theory will face essentially the same problem discussed in
this section.
Specifically, having S is grounded in believing that S is the most truth-conducive standard available – and clearly it is incoherent to both believe something and not believe it.19 With the belief about the truth-conduciveness of S in place, an agent clearly cannot rationally form beliefs using any standard other than S, even if the other standards are rational. This would be using a standard that she believes to be less likely to arrive at true beliefs, which is patently irrational.

It is clear, however, that this view cannot satisfy the applicability criterion. Few, if any, actual agents have any beliefs about epistemic standards at all, much less beliefs about the truth-conduciveness of any particular epistemic standard. Most agents have never even thought about the concept. Since real life agents are supposed to have epistemic standards, this theory cannot be correct.

Seeing this, we might want to retreat to a normative notion. We might think that an agent’s epistemic standard is the standard the agent is epistemically required to believe to be the most truth-conducive standard available.

If we take this route, however, what should we say about why agents have this epistemic requirement? One possibility is that it is like any other epistemic requirement. But recall that under the views in question, an agent is required to form a belief just in case the agent’s epistemic standard, coupled with her evidence, advises forming that belief. But if this epistemic requirement is what determines which epistemic standard an agent has in the first place, the explanation is clearly circular. On the other hand, we might simply say that there is no further explanation for why agents are required to believe that some specific standard is especially truth-conducive. Such a theory, however, is rather empty. It is, in effect, claiming that there is

---

19 In fact, this doesn’t seem just incoherent, but perhaps impossible. This may actually be a further weakness of the belief-theory: it satisfies the normative criterion by making this seemingly possible belief impossible.
nothing about an epistemic agent that makes her have one epistemic standard rather than another – it is simply a brute fact.

In addition, retreating to a normative notion fails on other grounds. Recall that we need a theory which allows for the possibility of agents having irrational epistemic standards. Under the current theory, however, an agent with an irrational epistemic standard is an agent who is required to believe that her irrational standard is the most truth-conducive one available. But this means that an agent with an irrational standard that includes affirming the consequent, for example, would be epistemically required to believe that affirming the consequent tends to lead to true beliefs. This is clearly not anything that an agent can be required to believe. Thus, such a theory cannot plausibly allow for the existence of irrational epistemic standards. So it would not satisfy the applicability criterion anyway.

6. The Disposition Theory of Standard-Possession

6.1. Dispositions and Performance Errors

If thinking about an agent’s beliefs does not get us a good theory of standard-possession, then perhaps the answer lies in considering an agent’s dispositions. There are different dispositions we might look to when constructing a theory of this type. For example, a strain of Bayesian thinking analyzes credences in terms of dispositions to place certain bets. Since under subjective Bayesianism, epistemic standards are simply encoded in an agent’s prior credences, this also serves as a theory of standard-possession. Alternatively, we might look toward what an agent is disposed to believe about how to form beliefs – her higher-order belief

20 Note that if some dispositional theory of beliefs is correct, then the belief theory of standard-possession is a version of the dispositional theory of standard-possession.
21 For a classic presentation of this type of view, see (Ramsey 2010).
dispositions. Or perhaps we should look toward an agent’s dispositions in some specific counterfactual situations which reveal their “deepest” epistemic selves.

Here, however, we will focus our discussion the simplest version of a dispositional theory – one which looks toward which first-order beliefs an agent is disposed to form. It will turn out that the main problem with this theory will afflict other dispositional theories of belief formation as well.

The idea that an agent’s epistemic standard is fixed by her belief-forming dispositions is not only simple, it is intuitively plausible. After all, an agent’s epistemic standard is the one that she uses – and this will presumably show up in the way that she forms beliefs. It also seems to have a good chance of avoiding the problem that plagued our last attempt. By tying an agent’s standard to what she would believe, we have found a criterion that clearly applies to actual non-ideal agents in the real world.

Unfortunately, if we simply say that an agent’s epistemic standard is the one that maps onto her belief-forming dispositions, we end up with a different problem. This theory has the consequence that all agents live up to their own epistemic standards far too much of the time. This is because under this theory, it is impossible for an agent to have dispositions that fail to live up to her own standard – after all, those dispositions determine which standard the agent has in the first place. Intuitively, however, agents can be irrational in this way. Consider the case of implicit bias. Many actual agents are often disposed to form beliefs in a sexist or racist manner. This seems to be due to some kind of unconscious psychological mechanism. Many of these agents would explicitly reject racist or sexist methods of belief-formation. They would be inclined to take back any beliefs formed in such a manner if their biases are disclosed. And they would think that such patterns of belief-formation are irrational and inaccurate. All this is
sufficient reason to think that these agents don’t necessarily have sexist or racist standards in all cases. Instead these agents consistently fail to live up to their own standards.\textsuperscript{22}

\textbf{6.2. Idealized Dispositions}

The solution again seems to be to idealize. There are at least two different ways of doing this. One option is to idealize the agent. To idealize a given agent, we need to think about the nature of the idealized thinker. This thinker, among other qualities, has infinite time and conceptual resources to consider all her evidence, has no motivations other than the truth, and has no cognitive limitations. The idea is to imagine what an agent will be like if she has all of the epistemically “good” qualities. After finding the agent’s ideal counterpart, we see which epistemic standard is encoded in the ideal agent’s dispositions. That, we then claim, is the epistemic standard possessed by the non-ideal agent.

Unfortunately, this kind of idealization again runs into problems with the applicability criterion. It hard to see how any agents could have irrational standards under this theory, since idealized thinkers would not have irrational belief dispositions. For example, it is implausible that there is a possible idealized thinker who affirms the consequent, since being able to see logical entailments is part of our conception of an idealized thinker. So no agent’s idealized counterpart will affirm the consequent. Which means no agent can have an epistemic standard which affirms the consequent. Thus, this kind of idealization seems ill-suited for the purpose of a theory of standard-possession.

Alternatively, instead of idealizing the agent, we can try to idealize the agent’s belief-forming dispositions themselves. To do this, we can look toward an analogy with Humean

\textsuperscript{22} (Kripke 1982, pp. 28–30) identifies similar problems with dispositional theories of rule-following.
conceptions of laws of nature. Roughly, Humeans view the universe as, in the first instance, a vast mosaic of particular facts (properties spread out over a space-time manifold, perhaps). Laws of nature are those axioms which form a deductive system which best describes the Humean mosaic – has the best tradeoff between the theoretic virtues of fit and simplicity.

Analogously, for the purposes of identifying an agent’s “laws of thought” we might first consider the “mosaic” of belief-forming dispositions (doxastic states spread out over a manifold of possible evidential states, perhaps). We then find a system that describes the mosaic with the best tradeoff of fit and simplicity. That system is the agent’s epistemic standard.  

This strategy seems to work in many cases, correctly identifying the performance errors. For example, suppose an agent generally reasons in a way that is deductively valid, but affirms the consequent in some very particular circumstances. We can choose between two epistemic standards – one that just follows the rules of deduction, and another that mostly follows the rules of deduction but allows exceptions in very specific cases. Plausibly, the first standard is preferable, since it is much simpler than the second one, and is only slightly worse at fitting the mosaic. Thus, the instances of affirming the consequent are counted as performance errors. On the other hand, if the agent is constantly disposed to affirm the consequent, then the second standard might become preferable. Despite being less simple, it fits the mosaic much better than the first. So here, the agent has an epistemic standard which endorses affirming the consequent. This lines up with our intuitions about the cases.

That being said, it is not clear that this strategy will get all the cases right. Some performance errors do follow easily discernable patterns. Implicit biases are cases of this.

---

23 See (Lewis 1983, pp. 365–8) for a presentation of the Humean theory. This is, of course, not the only way to idealize belief-forming dispositions. Notably, in the same paper (pp. 375-6) Lewis presents a theory of rule-following that can be adapted for this purpose. This method faces the same problems as the Humean method, so looking at the failures of the Humean method will be enough for the purposes of this paper.
Suppose an agent consistently undervalues women in a wide variety of contexts. One possible epistemic standard is one that judges women and men equally. The other is one that judges women more harshly than men. Even if the first standard is simpler than the second, it is not clear that this consideration will outweigh how much better the second standard does with regards to fit. So this theory might not rule the implicit bias as a performance error – which it intuitively is.

Still, this type of idealization is probably our best bet when trying to construct an idealized disposition theory of standard-possession. So let us just assume that it will deliver the intuitive results, correctly ruling on whether irrational beliefs are performance errors or successful implementations of irrational standards. In this way, the idealization procedure satisfies the applicability criterion. The view we end up with is one where a certain rough pattern in an agent’s mosaic of belief-forming dispositions is “highlighted” as the agent’s epistemic standard. The next question is whether such a view can satisfy the normative criterion.

6.3. Dispositions and the Normative Criterion

Under this theory, is it plausible that an agent is epistemically required to form beliefs in accordance to her own epistemic standard, rather than any other rational standard? There is at least an initially intuitive story that defenders of the ESV can tell here. Again, the key is to focus on the Humean idealization. The Humean procedure outputs a pattern in an agent’s disposition mosaic that has the best trade-off between simplicity and fit. It is this pattern that an agent is actually required to conform to – not the agent’s actual dispositions.24 Actual dispositions of an

24 This is important because of (Kripke 1982)’s famous complaint that an agent’s dispositions at most tells us what she will do, not what she should do. The mere fact that an agent is disposed to form a certain belief given a certain body of evidence cannot settle the question of whether that belief is rational for her. Thus, if we had to say that an agent’s standards are just her dispositions, we would be unable to answer even the most general normative question,
agent are only relevant insofar as they determine fit – in essence, providing a “distance” measure for how close an agent is to a particular pattern. It seems intuitive to think that the Humean pattern is somehow better than an agent’s actual dispositions, epistemically speaking. Perhaps they are even epistemically ideal. If this can be made to work, then we have a theory whereby agents are required to form beliefs in accordance with the epistemic ideal with which they are “closest.” At least on first blush, this seems plausible.

We have to be careful here. Remember that Humean patterns are not defined by the dispositions that the agent as an ideal thinker would have. That way of idealization runs into problems with the applicability criterion. So what makes these patterns epistemically better cannot be any direct connection with our conception of the ideal agent. Thus, the normative force we are seeking cannot come from any thought that agents should emulate ideal versions of themselves. Instead, it seems that defenders of the ESV need to claim that somehow simple patterns of belief formation are epistemically better than less simple ones. It is beyond the scope of this paper to explore if such a claim is plausible. We will merely note that more work is needed to explain why the Humean idealization procedure epistemically improves upon an agent’s belief-forming dispositions.

In any case, even if this problem is solved satisfactorily, we have not yet begun to address the normative criterion. To see this, let us grant the defender of the ESV as much as possible. Let us assume that the Humean pattern of belief formation which constitutes an agent’s epistemic

\[\text{viz. why should an agent believe in accordance with any epistemic standard at all? However, even if we can answer Kripke’s challenge for epistemic standards, and we understand why we should follow rational standards (rather than irrational ones), there is still another explanatory challenge: why should we follow our epistemic standard rather than any other rational standard? That is, the normative criterion is still not necessarily satisfied.}\]

\[25\text{ If we wanted to continue to follow Lewis, for example, we might make a distinction between “natural” and “unnatural” predicates, and then claim that simpler patterns are more natural, and therefore, epistemically better. It is unclear whether this can be made to work in the end. Even if it can, it may require metaphysical or epistemic baggage that will ultimately prove too burdensome.}\]
standard is epistemically ideal. So following a Humean pattern is a necessary condition for rational perfection. Remember, however, that the ESV is a form of permissivism. So there must be multiple patterns which are epistemically ideal. Even if we can explain why agents should prefer Humean patterns to non-Humean patterns, there is another explanation needed – why should an agent prefer any particular Humean pattern over others? Specifically, the present theory states that when choosing between two equally good patterns to follow for some particular belief, an agent is required to follow the pattern that fits her actual dispositions the best – the pattern that she is “closest” to. This goes unexplained. Forming the closer belief might be easier or less upsetting than the further away belief. All this might show that this is an imprudent thing to do. But if an agent chooses the further belief, why is that any epistemic mark against her?

To see the problem more clearly, consider an analogy. Assume, as is plausible, that there is more than one ideal set of tennis dispositions, as there is more than one ideal tennis “playstyle.” If this is true, then there is a natural way to advise a tennis player. We look at which playstyle a player possesses – which involves looking at a player’s tennis dispositions, and figuring out which idealized playstyle her dispositions are “closest” to. And then we say “perfect that playstyle.” If she successfully does this, we would think that she is playing perfect tennis.

Imagine a player who almost perfects her playstyle. But suppose that, for whatever reason, she does not completely embrace it. In some very specific situations, the player plays differently. Let’s say, for example, that this happens every leap day. On leap day, she behaves as if she was perfectly embracing a different playstyle. This may be extremely difficult, since it involves changing her style of play in very small set of cases. But suppose that, against the odds, she is successful at perfectly adopting a different playstyle for one day every four years. Intuitively, it does not seem that there is anything wrong with her tennis playing. She may be
acting imprudently, since she seems to be putting undue burden on herself, but she is not any less perfect as a tennis player.

It is worth noting that just as in the belief case, the problem for tennis flows directly from our assumption of permissivism. If there is only one ideal playstyle, proclaimed by fiat of the tennis gods, then there is no question as to why a player needs to play according to that playstyle. The problem arises when multiple playstyles are all good from the God’s-eye point of view. Then, it seems, a player can arbitrarily pick and choose from among them. Of course, this is less absurd in the case of tennis than the case of rational belief.

Thus, we have no explanation for why an agent must believe according to her epistemic standard rather than any other rational standard. So the normative criterion is not satisfied by the dispositional theory of standard-possession.

7. No Explanation Needed?

After seeing two broad families of theories fail to satisfy both criteria, one might become suspicious of the criteria. Maybe the problem is that we are asking too much from our theories of standard-possession. Specifically, maybe the normative criterion is an unfair burden. After we understand what it means for an agent to possess a certain standard, then perhaps it is simply implausible that agents are permitted to form their beliefs in accordance with any other standard. Perhaps no deeper explanation is needed.

To see why this type of skeptical response might be plausible, we can look at an analogy. Consider the case of decision theory, understood as a theory of prudential rationality. Under this theory, what an agent should do, prudentially speaking, is a function of her beliefs (at least if they are rational) and her desires. It does not seem mysterious why decision theory takes the
agent’s own beliefs and desires in account rather than anyone else’s. No additional explanation seems needed. If this is the case for decision theory, maybe we can say something analogous for the case of epistemic standards.\textsuperscript{26}

While it may seem like that there is no analogue of the normative constraint for decision theory, this is only because the target explanation is so deeply entrenched in the concept of prudential rationality that such an explanation is hard to see. Seeing the explanation will help elucidate why the satisfaction of the normative criterion is in fact necessary for a plausible version of the ESV.

Let us start with desires – why does decision theory take an agent’s own desires into account, rather than someone else’s desires? After all, it is natural to think that all desires are equally rational, as far as prudential rationality is concerned. So from the point of view of any agent, isn’t any set of desires as good as any other? To answer this question, we need to look no further than the concept of prudential rationality itself. An agent is being prudentially rational, it seems, just in case she is doing what she should expect to best satisfy her own desires. That is, it is a conceptual truth about prudential rationality that it is directed toward each agent’s own desires – in this sense prudential rationality is agent-relative. This is why decision theory takes an agent’s own desires into account. So while it is true that this fact is not mysterious, it is only because the explanation is obvious, being built into the concept of prudential rationality.

Suppose we wanted to appeal to an analogous explanation for the case of epistemic standards. This would involve claiming that epistemic rationality just is a matter of forming beliefs in accordance to one’s standards – it is part of the concept of rationality. Even if such a move was plausible, it is still the case that not just any theory of standard-possession would

\textsuperscript{26} I would like to thank Christopher Meacham for pressing me on this point, and for providing the decision theory analogy.
satisfy the needs of the ESV. Again, we can see this if we combine this view with the hair color theory of standard-possession. We end up with a highly implausible picture of rationality – forming rational beliefs just is a matter of forming beliefs in accordance with the standard associated with your hair color. This is a kind of “rationality” that no one could possibly care about.

Alternatively, the defender of the ESV might posit a different agent-relative factor toward which epistemic rationality is directed. Then, if she can show that such a factor is also correlated with one’s epistemic standard, then the normative criterion might be satisfied. We might say, for example, that epistemic rationality is directed toward the satisfaction of an agent’s desires. Then, if we can show that forming beliefs in accordance to one’s epistemic standards is conducive to desire satisfaction, then the defender of the ESV has the explanation she needs. But note that not just any theory of standard-possession could work here. Again, if we couple the desire satisfaction conception of rationality with the hair color theory of standard-possession, we would not be able to satisfy the normative criterion – people’s desires do not generally correlate with their hair color. The upshot of all of this is that if the defender of the ESV wanted to avail herself of this kind of explanation, the details of her theory of standard-possession must allow it – and not just any theory will. And so, her theory must still satisfy the normative criterion.

Let us move on to the case of beliefs. Decision theory takes an agent’s own credences as input, rather than any other agent’s credences. An explanation for this which appeals to conceptual truths about prudential rationality might also work here. After all, prudential rationality answers the question of “what should I do?” And it might be a conceptual truth that what one should do is dependent upon her own view of what the world is like. If this is right, then the explanation for why decision theory takes into account an agent’s own desires, and why
an analogous explanation does not exist in the case of epistemic standards, carries over directly to the case of beliefs.

This being said, it is less obvious that the conceptual explanation works in the case of beliefs. After all, there are almost always going to be sets of credences that are more accurate than an agent’s own, even if they are perfectly rational. And since more accurate credences are more conducive to satisfying an agent’s desires, why doesn’t decision theory take the more accurate set of credences as its input? The answer, I take it, is because there is no way for an agent to know that any set of credences are more accurate than her own. Indeed, it would be irrational for an agent to believe this about any other set of credences. This would involve having a higher-order belief that one’s first-order credences do not maximize expected accuracy. This constitutes a type of irrational internal inconsistency. After all, if a rational agent did somehow find herself in such a situation, she would simply change her first-order credences to those which she thinks are more likely to be accurate. So upon considering any other set of credences, a rational agent will think that her own are closer to the truth – and therefore, better at getting her what she wants. Thus, prudential rationality takes into account an agent’s own beliefs because those are the beliefs a rational agent will take to be most expectedly accurate, and therefore, most conducive to satisfying the agent’s desires.

The fact that agents are rationally required to believe their own credences to be the most expectedly accurate will not, however, help us satisfy the normative criterion for epistemic standards. The rational constraint in question is a consistency requirement between an agent’s first-order beliefs and her higher-order beliefs. But none of this has anything to do with an agent’s epistemic standards, at least not directly. Without additional details about what it means for an agent to have a particular standard, it seems possible for an agent’s first-order and higher-
order beliefs to be completely consistent with each other and still fail to align with their own epistemic standard (while perhaps aligning with some other rational standard). A good theory of standard-possession needs to rule out this possibility. Thus, the normative constraint still stands as an adequacy condition on any good theory of standard-possession.

8. Conclusion – A Diagnosis

We have canvassed two different types of theories of standard-possession – which include all the extant theories – and found that none of them can serve the broader purposes of the ESV. We have also seen that denying the need to satisfy the normative criterion is implausible. This gives us a good reason to think that the ESV is in real trouble.

This being said, it would be nice if we could give some kind of diagnosis as to why all of these theories have failed. This would give us an in-principle reason to think that there are real obstacles to successfully implementing a theory of standard-possession within the ESV. We will conclude by attempting to do this. The failures of the theories we have considered reveal three tensions within the ESV – tensions that make it extremely difficult to implement any theory of standard-possession.

First, there is a tension in how much leeway defenders of the ESV actually think rationality gives epistemic agents. On one hand, the ESV says that there can be a wide variety of different epistemic standards that an agent can be rational in having. So from the third-person perspective there can be a wide variety of permitted responses to some bodies of evidence. Rationality thus seems quite permissive. On the other hand, from the point of view of any one epistemic agent, no bodies of evidence can permit her to respond in more than one way. She can only form the unique belief advised by her standards. From the first-person perspective,
rationality is actually very constraining. So a theory of standard-possession has to be liberal enough to allow that all sorts of standards are rational, but limiting enough to bind any one agent to a single standard. This is why the normative criterion is hard to satisfy.

Second, there is a tension in the way the ESV conceives of actual agents. On one hand, the ESV is optimistic about the rationality of actual agents. Actual agents, the ESV claims, are often rational. When they disagree, it is often because they have different, equally rational standards, rather than because anybody is making a mistake. And when agents have problems with implicit bias or making deductively invalid inferences these are often just performance errors, rather than revealing biased or illogical epistemic standards. On the other hand, however, the ESV needs to claim that actual irrational standards – probabilistically incoherent ones or counter-induction endorsing ones, for example – are at least possible. The ESV is thus very optimistic on one hand and very pessimistic on the other. So a successful theory needs to resolve this tension somehow. This makes satisfying the applicability criterion very difficult.

Finally, there is a tension between the two criteria as well. On one hand, we cannot idealize too much in fixing an agent’s standards. Otherwise, possessing irrational standards become metaphysically impossible. But idealization is required for explaining why an agent must adhere to her standards, rather than some other standard. After all, it is plausible to think that agents should believe in accordance to some ideal. Without idealization, it is harder to see the normative force of possessing a particular standard. So the two criteria pull in opposite directions.

The ESV is a promising and popular way of developing permissivism – a way that answers to two popular motivations which apply to all permissive theories. But in order to work, the ESV needs an account of a central notion it is built on: the notion of an agent’s possessing a
particular standard. Extant versions of the ESV have offered only sketches of such accounts, and
the sketches they have offered have not been plausible. We have seen how an acceptable
account must satisfy two criteria, and we have seen why meeting both criteria will be difficult.
So while we have seen no proof that a satisfactory account is impossible, we have seen that any
theory of standard-possession that has any hope of fulfilling the ambitions of the ESV has to deal
with an entire minefield of problems due to the central tensions within the view. Thus, there is a
serious obstacle to formulating a complete and satisfying version of the ESV.

9. Works Cited

Imprint, 11*(18).


Without Gods: Meditations on Atheism and the Secular* (pp. 194–214). OUP.


HOW SUPEREROGATION CAN SAVE INTRAPERSONAL PERMISSIVISM

Abstract: Rationality is intrapersonally permissive just in case there are multiple doxastic states that one agent may be rational in holding at a given time, given some body of evidence. One way for intrapersonal permissivism to be true is if there are epistemic supererogatory beliefs – beliefs that go beyond the call of epistemic duty. Despite this, there has been almost no discussion of epistemic supererogation in the permissivism literature. In this paper, I show that this is a mistake. I do this by arguing that the most popular ways of responding to one of the major obstacles to any intrapersonally permissive theory – the arbitrariness objection due to Roger White – all fall prey to the same problem. This problem is most naturally solved by positing a category of epistemically supererogatory belief. So intrapersonal epistemic permissivists should embrace epistemic supererogation.

1. Introduction

Rationality constrains the beliefs we are allowed to have – this much is obvious. In any situation, there are certain beliefs that are simply irrational to hold. What is less obvious is how much rationality constrains our beliefs. How strict is rationality? Is the rational agent chained by the dictates of reason, such that her opinions are shaped only by contingencies of her situation? Many writers intuitively think (and perhaps many more hope), that this is not true. That is, maybe even the most zealous follower of reason is allowed some latitude in her beliefs. Maybe rationality allows us some leeway in what we ultimately believe. Perhaps rationality is “permissive.”

Here is one way that rationality might be permissive. Even if there is always one unique belief that is maximally rational to hold in any given situation, perhaps agents are not irrational if
they fail to hold this belief. Sometimes, there are beliefs that are rationally permissible to hold even if they are not maximally rational. Maybe rationality sometimes cuts us slack, letting it go even if we don’t always do the optimal thing. For rationality, sometimes good enough is good enough. What I am suggesting is that there may be epistemically supererogatory beliefs – beliefs that go beyond the call of epistemic duty. This opens up the conceptual space for beliefs that might meet epistemic duty – are rationally permissible to form – even though there are better beliefs available. There is, of course, a large and varied literature on moral supererogation. But there has been very little written on the epistemic counterpart.  

We should find this surprising. At the very least, there are many seeming parallels between ethics and epistemology, the two major normative philosophical disciplines. Investigating these parallels is a good strategy to make progress in both fields, even if, on the final analysis, the seeming parallels turn out to be misleading. This in itself makes it surprising that epistemic supererogation is rarely discussed. It is even more surprising that discussions of epistemic permissivism rarely talk about supererogation, even though epistemic supererogation is clearly one way for rationality to be permissive. In this paper, I will argue that this is a mistake. Indeed, defenders of a certain kind of permissivism should embrace epistemic supererogation.

Before we begin, some clarifying remarks are in order. Permissivism is traditionally conceived of as the denial of a thesis known as “uniqueness.” In this paper, we will follow this tradition. Here is a fairly orthodox formulation of the thesis, which is more than adequate for our

---

27 One exception is Hedberg (2014), although that paper is about epistemically supererogatory actions (such as gathering additional evidence or double checking past evidence), whereas I am interested in epistemically supererogatory doxastic states.
28 This point is emphasized by Berker (2013).
29 Though, as we will see, some proposals can be interpreted as forms of epistemic supererogation. Podgorski (2015) perhaps comes closest, when it says that you can be “doing better than you need” (§13) when you undergo some non-required epistemic processes.
purposes:

**Uniqueness:** Necessarily, for any total body of evidence E, and proposition P, there is at most one doxastic attitude to take toward P that is consistent with being rational and having E.\(^{30}\)

In addition, most permissivists defend an especially strong version of the view, such that there are situations which are permissive between believing some P and believing ~P (as opposed to believing P and suspending judgment on P, for example).\(^{31}\) In this paper, the term “permissivism” will be referring to the strong version of the view.

Permissivism can be further delineated into two more specific claims, which we will call interpersonal permissivism and intrapersonal permissivism.\(^{32}\) Roughly, interpersonal permissivism says that two (or more) persons may have different doxastic responses to the same body of evidence and both be rational. Intrapersonal permissivism, on the other hand, says that the *same* person may have one of two (or more) doxastic responses to the same body of evidence and be rational in either response. More specifically, in this paper intrapersonal permissivism will refer to the claim that there can be a point in time where a single agent, with some body of evidence, can form either of two (or more) incompatible doxastic states and be rational no matter which way she goes.\(^{33}\) In this paper we will be exclusively interested in the prospects for

---

\(^{30}\) This formulation is slightly modified from Schoenfield (2014), 3. In this paper, I will assume that an agent is rational insofar as she has epistemically justified beliefs. Accordingly, I will use the phrases “rational belief” and “justified belief” interchangeably.

\(^{31}\) In this paper, I will be talking in terms of all-or-nothing beliefs, although I suspect everything I say can be translated into talk of credences.

\(^{32}\) This terminology is from Kelly (2013).

\(^{33}\) The point of using the terminology in this way is to distinguish intrapersonal permissivism from what we might call “possible worlds permissivism,” such that the same person, in two different possible worlds, might be rational in having different beliefs in those two worlds even with the same evidence. This view seems weaker than intrapersonal permissivism (in my sense), and is normally held by interpersonal permissivists who do not want to commit themselves to full on intrapersonal permissivism. This is because interpersonal permissivists posit some non-evidential fact about persons that make certain beliefs rational for them. These non-evidential facts can vary across different people, which can also change what is rational for an agent to believe, even with the same body of evidence. But these facts are generally not thought of as essential features of any single person – which means that they can also vary across possible worlds. So the same agent can vary with respect to this non-evidential fact across
intrapersonal permissivism. Thus, the thesis should be understood as limited in this way: intrapersonal permissivists should embrace epistemic supererogation.\textsuperscript{34}

It should also be noted that the word “rational” in the definition of uniqueness should not be understood to mean “maximally rational.” This is because we want theories of epistemic supererogation which posit one maximally rational doxastic response to any body of evidence to count as permissive, as long as they also posit some other permitted (though worse) doxastic response. Perhaps because writers typically do not countenance the possibility of epistemic supererogation, the distinction between “maximally rational” and merely “rational” (or “as rational as one is required to be”) is rarely explicitly discussed when attempting to define permissivism.\textsuperscript{35} It is possible that after we make the distinction explicit, some defenders of uniqueness will realize that “maximal rationality” is what they were interested in all along. Which means that a theory of epistemic supererogation will be consistent with what these philosophers want to call “uniqueness.” Still, enough theorists will agree with our definition of uniqueness for us to reasonably count epistemic supererogation as a form of rational permissivism (as it intuitively is). As we will see, however, there is a sense in which a theory of epistemic supererogation is “closer” to uniqueness than other permissive theories, since it is

\textsuperscript{34} Though this paper does not argue for the existence of epistemic supererogation for anyone who is not already committed to intrapersonal permissivism, I do think that epistemic supererogation is independently plausible. See (Li 2017)

\textsuperscript{35} One exception is Christensen (2007) and (2013), who makes his formulation of uniqueness thesis explicitly about maximal rationality. As evidence for the claim that most philosophers are simply not thinking of this distinction, consider the following (incomplete) sampling of how the literature talks about the uniqueness thesis. Some writers such as White (2005), Douven (2009), Cohen (2013), and Schoenfield (2012) simply talk about what is “rational,” while Feldman (2007) defines uniqueness in terms how many propositions the evidence “justifies.” White (2013) and Kelly (2013) define uniqueness in terms of how many doxastic states are “fully rational.” (To my ear, “full rationality” is ambiguous between “maximal rationality” and “as rational as one is required to be.”) Titelbaum and Koepc (ms) present different variations of the uniqueness thesis, with different versions alternatingly talking about what the evidence “justifies,” “confirms,” or “rationally permits.” Similarly, Ballantyne and Coffman (2011) use both the term “rational” and “justifies” in the same formulation.
consistent with a natural line of thought that, *prima facie*, seems to lead to uniqueness. This is another mark in favor of epistemic supererogation.

That being said, not just any theory of epistemic supererogation can serve as a plausible form of permissivism. A theory, for example, which draws the line of rational permissibility by considering everyday judgments of rationality would be ill-suited for the purpose. This is because even the most ardent defender of uniqueness will admit that in non-theoretical contexts we talk loosely about rationality. Ordinary language allows for deviations from maximum rationality that we nonetheless call “rational” because of the need to make epistemic appraisals of real agents – agents who rarely achieve maximally rational doxastic states. If this is the reason that epistemic supererogation exists, then the truth of uniqueness is not threatened. The type of permissivism that such a theory of supererogation represents would hold little theoretical interest, and the debate over uniqueness could easily be reinterpreted as a debate about maximal rationality.

Fortunately, a theory that relies on everyday ascriptions of rationality and the cognitive capacities of actual agents is not the only option available. To see this, we can examine the analogy with moral supererogation. Many defenders of moral supererogation maintain that there is still a bright line between permissible and impermissible actions. To say that an action is minimally permissible is not simply to say something about its goodness compared to other options – it is not simply saying, for example, that the action is not horribly bad. For even if some permissible actions are not maximally good, performing those actions still discharges one’s moral *duties* – and there is a real moral difference between doing one’s duty
and failing to do one’s duty.\textsuperscript{36} Thus, we can construct an analogous concept for the epistemic realm, such that rationally permissible doxastic states are epistemically different from the rationally impermissible ones in a theoretically interesting way, even if they are not maximally rational.

Detailed theories of epistemic supererogation can be developed in many different ways, although this paper will not rely on the details of any specific theory. I will briefly mention, however, a particular type of theory that I find most promising.\textsuperscript{37} Some theorists contend that moral supererogation happens because actions can be judged with regards to two different types of moral virtues or values.\textsuperscript{38} Roughly speaking, the idea is that one of these values (the value of justice) can require certain actions of agents while the other value (the value of beneficence) justifies certain actions as better than others. Thus, certain beneficent actions are morally supererogatory because the value of beneficence cannot generate moral requirements. This type of theory can serve as a model for epistemic supererogation if we can also find two different epistemic values, one of which plausibly generates rational requirements while the other generates rational justification. Such a theory, if successful, would serve as a type of rational permissivism.

In what follows, we will begin with some remarks on the intuitive plausibility of both intrapersonal permissivism and epistemic supererogation. Next, we will consider a general difficulty for intrapersonal permissivism – the “arbitrariness” objection. We will then examine

\textsuperscript{36} Indeed, in his seminal work on moral supererogation, Urmson (1958) explicitly distinguishes between an agent doing her duty (even in contexts where this is extremely difficult, such that ordinary persons would fail to do their duties) and going beyond the call of duty – that is, performing supererogatory actions.

\textsuperscript{37} See (Li 2017) which develops this type of theory in detail. It is also worth mentioning that though we have relied on the analogy between moral supererogation and rational supererogation here, the analogy eventually comes apart. Thus, we must be mindful of the disanalogies between the two normative realms, and show that they are not an obstacle to developing a plausible theory of epistemic supererogation.

\textsuperscript{38} Theories with this idea at their core have been proposed by Zimmerman (1993) and Dreier (2004), although here I am not ascribing the view to anybody in particular.
two different types of response to this objection. It will be shown that the two types of response both face the same serious problem as they stand. We will see that positing a category of epistemically supererogatory belief is the natural way to deal with this problem. So it will turn out that different paths to avoid the arbitrariness objection all lead to epistemic supererogation. This seems like good reason for intrapersonal permissivists to embrace epistemic supererogation.

2. Intuitive Considerations

Perhaps the most widely cited reason for thinking that rationality might be permissive is the intuitive existence of rational disagreement, even on the same evidence. Rosen (2001), for example, writes that when a jury is divided on a difficult case, or when paleontologists disagree about what killed the dinosaurs, it does not necessarily mean that someone has irrational beliefs. Kelly (2013) writes about a case where different agents disagree about how likely it is for a particular candidate to win the presidency before a close election. And Douven (2009) has argued that in some cases, scientists can rationally disagree about which theory best explains a body of evidence.

In these examples, we not only have intuitions about the existence of disagreement, but we clearly also have intuitions about what types of evidence engenders this type of disagreement. It is not a coincidence that these are cases where the evidence is extremely complex, multi-faceted, scarce, or fractured. Perhaps, then, certain bodies of evidence can have features that make them rationalize more than one doxastic response. When agents have this type of permissive evidence, reasonable disagreement can happen. If it’s the evidence that permits a range of rational responses, it’s plausible to think that even a single agent with this sort of evidence can go either of two different ways, and end up rational. Thus, we end up with a
natural way to start thinking about permissivism, and it is a way that is consistent with intrapersonal permissivism.

Though epistemic supererogation is rarely discussed, there is also something intuitive to be said for the idea. After all, many of the same considerations that motivate theorists to posit morally supererogatory actions also apply to certain beliefs. There exist beliefs that represent impressive feats of epistemic prowess, which seem to go beyond our epistemic requirements. Consider, for example, Einstein’s theory of special relativity, the proof of the existence of irrational numbers discovered by a member of Pythagoras’ school, or the cases solved by fictional detective Sherlock Holmes. At least intuitively, these feats seem like epistemic analogues of the supererogatory acts of moral saints and heroes.

What is important to notice is that the cases where intrapersonal permissivism seems most intuitively plausible are also the cases where epistemic supererogation seems most intuitively plausible – cases where the evidence is complex, multi-faceted, etc. The particular nature of these bodies of evidence seems to be at least a partial explanation for why agents are not rationally required to respond to the evidence in the absolutely best way. This means that intrapersonal permissivism and epistemic supererogation mutually reinforce the plausibility of each other, in addition to whatever plausibility each view enjoys independently. So before we even get into the details of seeing why intrapersonal permissivists should embrace epistemic supererogation, we can see that a defender of either view already has some reason to accept the other.

3. The Arbitrariness Objection

To begin our examination of intrapersonally permissive theories, we will first consider a
general problem for permissivism: the arbitrariness objection. If intrapersonal permissivism is true, then there is going to be some situation where an agent, given her evidence, can either believe P or believe ~P and be rational in either belief. Here, however, we can begin to see an objection forming. For if both P and ~P are rational to believe, then it seems that there is no reason to believe one over the other. Suppose that the agent knows all this about her epistemic situation. Then even from the agent’s point of view, choosing one of these beliefs over the other seems awfully arbitrary.

To make this point vivid, imagine a rational agent who knows she is in a situation that is permissive between a belief in P and a belief in ~P. She also has two pills – one which induces a belief in P, and one which induces a belief in ~P. She could think hard about her evidence, weigh the different sides, and come to a belief based on that evidence, or she could randomly pick one of the pills and induce a belief in herself. Either way, she will end up with a rational belief. So why is any method better than the other? And after she does form a belief, why not take the pill for the contradictory belief? Again, either way she will end up with a rational belief. This situation is clearly absurd. Yet it seems that permissivists are committed to its possibility.39

One way to understand the worry behind this objection is to notice that there must be some connection between rationality and truth. Generally, when presented with a body of evidence we are warranted in believing that forming a rational belief given our situation is a good way to get to the truth. When an agent learns that her evidence is permissive, however, it seems that she also learns that the connection between truth and rationality is severed. To see this, suppose the agent considers a belief based on her evidence. Suppose she also knows that the belief is rational. With only this knowledge, it does not seem like she has good reason to

39 Many variations of this example were first suggested by White in his (2005) and then his (2013). Other versions of this worry are presented by Christensen (2007) and Feldman (2007).
think that the belief is likely to be true. After all, two incompatible beliefs are both rational, and only one of them can be true. So forming a belief in accordance with the evidence seems about as good as flipping a coin.\textsuperscript{40} Thus, it seems that when she learns her evidence is permissive she also learns that her evidence won’t do her much good. So an agent who realizes she has permissive evidence in regards to some proposition must suspend judgment on that proposition.

Thus, if permissive situations are supposed to be situations where two different doxastic states are equally rational, then when agents realize this, there is going to be another doxastic state that is better than both – suspension of judgment. And if suspension of judgment is better than the “tied” doxastic states, then that is the rational doxastic state for the agent to form. Which means that the situation is actually not permissive at all.

The arbitrariness objection is a roadblock that all intrapersonal permissivists must deal with. In general, there are two ways that permissivists have responded to it. In what follows, we will examine both strategies, and see why they both need to be complemented with a conception of epistemic supererogation.

4. There Are No Cases of Known Permissivism

4.1. The View

Faced with the arbitrariness objection, one option for permissivists is to concede the basic line of reasoning. It \textit{would} be absurd if an agent realized that she was in a permissive situation. So maybe such a realization destroys the permissivism. But this by itself does not rule out the existence of permissive cases – only cases of known permissiveness. The intrapersonal

\textsuperscript{40} Maybe the agent has some reason other than the belief’s rationality to think that it is likely to be true. But it is on the defender of permissivism to say what this other reason is. Later on, we will see one attempt at meeting the challenge.
permissivist, then, can say that permissivism is true, but only in cases where the agent does not know she is in a permissive case.⁴¹

This option involves accepting the thesis that Cohen (2013) has called “doxastic uniqueness,” which says that an agent cannot rationally believe that there are two (or more) rational doxastic attitudes to take toward some proposition P, given total evidence E, while holding either doxastic attitude and having total evidence E.⁴² There are two (compatible) ways to flesh out the details of doxastic uniqueness. One way is to think of the two doxastic states as “tied” or equally rational. The idea is that in some cases, an agent is permitted to form either of two tied doxastic states, as long as she doesn’t know they are tied.

The other model of permissivism involves thinking that permissive cases can happen even if one of the two permitted doxastic states is better supported than the other. This happens when, for whatever reason, the agent does not know that the rationally better belief actually is better. If this is rationally permissible, then the agent is permitted to form either the better belief or the worse belief. But it also seems that if the agent finds out the details of her situation – namely that some particular belief is better than some other particular belief – she must form the better belief. So again, when the agent realizes that her situation is permissive, her situation is no longer permissive.

There are different ways that one might fill in the details when attempting to deploy this general strategy. For illustrative purposes, let us quickly examine three of these attempts. I will illustrate the proposals with examples where the two permitted beliefs are tied, although they can also produce examples where one belief is better than the other.

---

⁴¹ As we will see, multiple responses to the White paper can be considered versions of this view. White himself considers it briefly in his (2005). Brueckner and Bundy (2012) also explicitly discusses this general strategy of responding to White.
The first possibility is suggested by Douven (2009). Douven’s proposal relies on the natural thought that, in many cases, some body of evidence supports a belief in some proposition because that proposition is a good explanation of the evidence. So for agents to form beliefs rationally, they have to think about how to explain the evidence they have. Some explanations, however, are extremely difficult to come up with – perhaps requiring nothing short of a brilliant flash of insight. But, Douven suggests, rationality may not require agents to have flash of insights. If this is right, then we have the possibility of permissive situations.

Imagine, for example, that a scientist has a large and complex body of evidence to consider. She attempts to come up with a good explanation for her evidence. As it turns out, there are two different and incompatible scientific hypotheses which explain the evidence equally well. Furthermore, both explain the evidence extremely well. The first hypothesis implies P, while the other implies ~P. Both explanations, however, are extremely difficult to come up with – so difficult that rationality does not require the scientist to come up with either. In this situation, if the scientist comes up with the first explanation she is rational to believe P, but if she comes up with the second one she is rational to believe ~P. So this situation is permissive. However, if the scientist comes up with both explanations, she will realize that her evidence supports both beliefs equally, and she can no longer either believe P or believe ~P.

Another possibility, due to Rosa (2012), relies on the thought that sometimes bodies of evidence can be inconsistent. According to Rosa, agents can have incompatible beliefs that they do not notice. In some of these situations, an agent can be rational in holding the inconsistent beliefs, and use them as evidence to form further beliefs. The idea is that these

\footnote{Douven (2009) 351-2, which features a permissive situation that arises due to some belief being supported by a brilliant flash of insight. Douven’s situation is one where one belief is actually better supported than the other, but as already mentioned, I have adapted it to cases of “ties” for expository purposes.}

beliefs can constitute inconsistent bodies of evidence which might support both a belief in P and
a belief in ~P. For example, suppose that some agent has total evidence consisting of:

(1) P and Q
(2) if P then R
(3) ~Q or ~R

That is, the agent rationally has all three of these beliefs without realizing that they are
inconsistent. With this evidence, she might reason to “R” from (1) and (2). Or she might reason
to “~R” from (1) and (3). We might think that depending on which reasoning process she goes
through, the agent is permitted to either believe R or believe ~R. Of course, if she goes through
both reasoning processes, then she is not permitted to form either belief. Indeed, she needs to
rethink her evidence.

Finally, Podgorski (2015) has suggested that which beliefs are rational for an agent might
depend on how much of her evidence she takes into account.45 And we might think that
sometimes agents are not required to consider all of their evidence all the time. Suppose, for
example, that an agent, receives two tiny bits of evidence regarding P. The evidence is small
enough that it is overwhelmingly likely to make no difference as to whether P. Maybe in this
case, the agent can neglect to think about these bits of evidence without being irrational. Of
course, the agent can consider the pieces of evidence. Suppose that it just happens to be one of
those cases where the small bits of evidence do make a difference as to whether P or ~P is likely
to be true. In fact, if the agent just considers the first of the two bits of evidence, then she is
rationally permitted to believe P. Had she considered just the second bit, she would be rationally
permitted to believe ~P. So this is a case where the agent could go either way, depending on

45 See Podgorski (2015) §10 and §12. Podgorski’s examples are most naturally interpreted as cases where one belief
is actually epistemically superior to the other. Again, for ease of exposition, I have given a case where the two
beliefs are “tied.”
which tiny bits of evidence she considers. Of course, if the agent knew all the facts about her situation – if, for example, she knew what would happen if she considered each bit of evidence – then she clearly is not permitted to form either belief.

4.2. The Problem

Just looking at the different ways of fleshing out the view, we begin to see where the apparent problem lies. Each one of these permissive situations only happen when the agent seems to exhibit some epistemic failing. The situations arise because the agent misses out on some explanations, reasoning processes, or considerations of evidence. Can any of this actually be rationally permissible?

One might think not. After all, evidential support relations seem knowable a priori. They are not the type of things that one typically gets empirical evidence for. Even if one did get such evidence, without knowing what the evidence supports (that is, without knowing the evidential support relations), the evidence would be no use. So at some point, agents have to be able to learn about these relations a priori. Thus, an agent with evidence that would permit believing either P or believing ~P should be able to tell that this is the case. But this just means that she is able to tell that her evidence is permissive a priori. Thus, the agent with permissive evidence without realizing it is permissive is simply not accessing all the evidential support relations available to her. One might object that this in itself is less than rational. If that is right, then any rational agent who knows all the support relations must suspend judgment. So no agent can actually be rational in either believing P or believing ~P, no matter what her evidence is like. In short, uniqueness turns out to be true.

More generally, the apparent problem for this view is that for an agent to be in a
position where she would be rational to believe P and also rational to believe ~P, the agent must be in a certain state of ignorance. This type of ignorance, however, is not due to some lack of empirical evidence on the agent’s part. Clearly, agents can figure out what their evidence supports by reflection – so this knowledge is available a priori. The agent could have conquered her ignorance by better thinking alone. The agent thus seems epistemically culpable for her ignorance, or so one might argue. At the very least, it seems that an ideally rational agent would not be so ignorant. Since it is intuitive to think that what an ideally rational agent would do is what we would be rational in doing, one might think that this strategy cannot really get us a permissive theory.

5. Epistemic Standards and Accuracy

5.1. The View

There is a different way to respond to the arbitrariness problem. This proposal employs the concept of “epistemic standards,” which we can think of as functions from bodies of evidence to doxastic states. Different standards encode different ways to respond to bodies of evidence. The general idea is that there is more than one standard that is rationally permitted. Each epistemic standard only advises a single attitude toward some proposition given a body of evidence. Different standards, however, might disagree about what this attitude is for some bodies of evidence. So there can be bodies of evidence where one rationally permitted epistemic standard will advise believing P, while another standard will advise believing ~P.

This type of view can avoid the arbitrariness problem because it allows truth and rationality to come apart, from the point of view of any individual agent. This is because agents identify with the particular epistemic standard that they rationally believe to be the most reliable
standard – the one most likely to get true beliefs.\textsuperscript{46} From an agent’s own point of view, other agents using different standards may be rational, and the beliefs they end up forming in accordance with those standards are rational \textit{for them}, but they are less likely to be true than the agent’s own beliefs.\textsuperscript{47} Thus, a rational agent can know that some situations are permissive. These are just the situations where her own standard disagrees with some other permissible standard, given her evidence. But choosing the belief her own standard advises isn’t arbitrary for that agent, since she also thinks it is more likely to be true than the alternative.

This way of thinking, however, does not seem to result in intrapersonal permissivism. Since each epistemic standard only outputs one attitude toward P, intrapersonal permissivism is only true if agents are permitted to use more than one standard. It does not seem, however, that agents are ever permitted to do this. To see this, consider what agents believe about the reliability of their own standards. If agents rationally believe that their favorite standard is the most reliable one, then they cannot be rational in switching to what they think of as a less reliable standard. If they believe that their standard is less reliable than some other standard, then they were not rational in using that standard in the first place. Finally, if agents believe that all the standards are similarly reliable, we seem to run into the arbitrariness objection again. This is, at least, a natural line of thought – and the reason this strategy has mostly been seen as a path toward interpersonal, rather than intrapersonal, permissivism.

\textsuperscript{46} Schoenfield (2014) pp. 7 adopts this view of what it means to have a standard. Elga (ms) fn. 3 espouses a similar view. Part of the idea is that an agent’s epistemic standards, if they are rational, must be “immodest.” An agent’s epistemic standard is immodest just in case it advises beliefs that maximize \textit{expected accuracy} from the agent’s own point of view, compared to the beliefs advised by any rival standard. Expected accuracy is a measure of how close an agent can expect some set of beliefs to be to the truth. Though this notion, favored by Bayesians, is usually defined in terms of credences, it is natural enough to discuss an analogous notion for all-or-nothing beliefs. See Lewis (1971) and Moss (2011) for examples of how to understand immodesty in terms of credences and the notion of expected accuracy. See Horowitz (2013) for a further discussion on immodesty as it relates to epistemic standards.

\textsuperscript{47} See Schoenfield (2014) 8-9 for a good presentation of this type of view. Subjective Bayesians also famously embrace this type of view.
There is perhaps one way of avoiding this line of thought. Titelbaum and Kopec (ms.) offer a response to the arbitrariness argument which denies that once we realize some body of evidence is permissive, we have no reason to think that forming a belief on its basis is likely to get us a true belief. Their view allows us to both say that rational epistemic standards are equally reliable at outputting true beliefs and that all standards are more likely to output true beliefs than false beliefs, including in permissive cases, where different standards output opposite beliefs.

To illustrate the proposal, let us focus on a toy example. Suppose that there are 100 rationally acceptable epistemic standards and 100 bodies of permissive evidence. For each body of permissive evidence, 90 of the acceptable standards output the true belief, and 10 of them output the false belief. But for any acceptable epistemic standard, it outputs the true belief on 90 different bodies of permissive evidence, and the false belief on 10 bodies of permissive evidence. Thus, even an agent who knows everything about the situation, including the fact that her evidence is permissive, can use her favorite rational standard and still be fairly sure (in fact, 0.9 sure) that the belief she ends up with is true. So an agent can choose any rationally acceptable standard she wants as her favorite. Whatever she chooses, the beliefs she forms with it will very likely be true – including the beliefs she forms based on evidence she knows to be permissive. Thus, it seems that we have a view allowing us to avoid the arbitrariness objection while also allowing that all rationally acceptable standards are equally reliable – giving us the intrapersonal permissivism we were looking for.

48 The view discussed here is especially inspired by §4 of their paper. Titelbaum and Kopec are not explicit on whether they mean their view to be a version of intrapersonal permissivism or not. So we can take the present view to be a version of intrapersonal permissivism inspired by their idea, which seems (at least initially) to be plausible.

49 Titelbaum and Kopec 21
5.2. The Problem

Titelbaum and Kopec argue that thinking about permissivism in this way affords us a nice way to maintain Conciliationism in the face of peer disagreement. If one of these agents ran into her equally rational friend and found out that they disagreed about one of these propositions, even though they had the same evidence, then she is no longer rational in maintaining her belief. This is true even if the disagreement is because the agents are using different standards (and everyone knows this). This is because after she learns of the disagreement, she has no more reason to think it is her favorite standard that is the one getting it right. It could just as easily be her friend’s. For example, suppose some agent rationally believes P, and finds out that her friend rationally believes ~P, based on the same evidence. She now knows one of them is in the minority with the wrong belief, but she has no idea which one. So she must give up her belief and suspend judgment. So her new situation, which includes the evidence of disagreement, is not permissive. Instead, she must conciliate.

While Titelbaum and Kopec see this as an advantage of their view, it also reveals a deeper difficulty. Let us ask whether the agent actually had to meet her friend in order to know that she would believe ~P. If she just knew what standard her friend was using, she could have figured out what her friend would believe (given that her friend didn’t make a mistake). And in general, the agent should be able to figure out what a person would believe, given some specific epistemic standard. Indeed, if she knew all 100 rational epistemic standards, then she could see which belief is in the majority for each body of potentially permissive evidence. Surely there is

---

50 Conciliationism refers to a broad family of views, according to which, a person should, upon discovering disagreement of a certain sort (normally disagreement about some proposition with someone with the same evidence and of similar intellectual prowess) revise her opinion in the direction of her disagreer. Titelbaum and Kopec consider the view’s compatibility with Conciliationism to be a nice feature of their view.

51 Titelbaum and Kopec (ms) 24-6.
nothing in principle stopping her from doing this. Presumably, coming up with standards, figuring out which ones are rational, and determining which belief they recommend in her specific situation, is not something that requires empirical evidence. It is an *a priori* matter. But since the majority of the equally reliable standards always advise the true belief, if the agent did this she would have the true belief every time. So this “meta-standard” of consulting all 100 rational standards and going with the majority is much more reliable than any one standard, even from the agent’s own point of view. One might object that the agent could not be rational in sticking to her favorite standard when she knows that all this is the case. And of course, all agents with the same knowledge of their situations should also use the meta-standard for the exact same reason. But this means all agents will be rationally required to come to the same beliefs, even given permissive evidence. In other words, it is hard to see how any of these bodies of evidence are permissive at all.

Once again, the apparent problem is that the purported permissive situations are ones where the agent is in a certain state of ignorance – in this case, ignorance about which standards are rational and what beliefs they would recommend. And again, since this ignorance is avoidable *a priori*, it does not seem to be a rational state to be in. If all this is right, then one might worry that we do not end up with a permissive theory at all.

### 6. A Diagnosis

We have seen two different attempts to construct intrapersonal permissive theories fall prey to essentially the same problem. In both cases, the purported situations in which an agent is permitted to either believe P or believe ¬P only occur when the agents are in a certain state of ignorance. Furthermore, they seem to be states of ignorance that are *a priori* preventable. Since
a defender of uniqueness could reasonably argue that it always seems irrational to be in a state of *a priori* preventable ignorance, these do not seem like situations that tell against the uniqueness thesis.

More abstractly, a uniqueness defender might argue that for any body of evidence, the possible doxastic responses to that body of evidence can be ranked according to their rationality. And this is a ranking that is determinable *a priori* – it is simply a matter of thinking up possible doxastic responses, figuring out how much they are supported by some body of evidence, and comparing their level of support. So every time an agent gets some evidence, she can just consult the ranking on that evidence. Once she has the ranking, it seems that she is rationally required to form the doxastic state at the top. After all, rationality is a guide to the truth. So, for any proposition P, if believing P is in the highest ranked doxastic state, then chances are P is true. And forgoing beliefs that are more likely to be true for beliefs that are less likely to be true is clearly irrational.52

For intrapersonal permissivism to be plausible, an alternative model must be offered. One possibility – that sometimes options are “tied” at the top of the ranking – has been blocked by the arbitrariness objection. In trying to escape this objection, both strategies we have canvassed rely on cases where the agent does not go through the entire procedure of constructing the ranking and finding the top doxastic state. In the first type of theory, the agent is in a permissive situation when she either does not know that her belief is tied with another one, or she does not know that another belief is ranked higher than her own. Under the Titelbaum and Kopec style view, the agent simply does not proceed to construct the ideal ranking, instead choosing to use a single epistemic standard and construct a non-ideal (but still fairly accurate)

52 Indeed, this might even be impossible, since it will likely involve believing P while also believing ~P is more likely to be true. This is dangerously close to both believing P and disbelieving P.
ranking. Without some explanation of how these situations are rationally permissible, both theories run into trouble.

7. Embracing Supererogation

At this point, it is hopefully becoming clear why embracing supererogation is a good way forward for epistemic permissivism. The belief which results from coming up with the complete ranking and choosing the top doxastic state is clearly the maximally rational doxastic state for an agent to form. If there cannot be ties at the top of such rankings, then there is only one maximally rational doxastic state for any body of evidence. As long as epistemic agents are required to form the maximally rational doxastic state, uniqueness will be true. The only way out, then, is to claim that sometimes agents are not required to do what is maximally rational. This means opening the way for epistemic supererogation.

Let us see more concretely how epistemic supererogation will help. Borrowing from the literature on moral supererogation, we can work with a somewhat bare definition of what it means for a belief to be supererogatory. Namely, a belief is epistemically supererogatory just in case it is (1) not rationally required, (2) rationally permissible, and (3) rationally better than some alternative belief that is rationally permissible.53

There are two ways we can employ epistemic supererogation in order to get us a theory of epistemic permissivism. We might say that, even after agents come up with the complete ranking of possible doxastic responses, agents are permitted to pick one that is not at the top. This means that there are some beliefs which are less than maximally rational, yet are permitted. So the maximally rational belief is (1) not required, (2) rationally permissible, and (3) rationally better than some other belief.

---

53 Plausibly, criterion (2) is redundant given criterion (3), since any belief that is rationally better than a permissible belief is itself permissible.
better than some alternative permitted belief (namely, the permitted non-maximally rational beliefs). However, for reasons already mentioned, I think it is implausible that agents can rationally form a belief that they know to be less rational (and therefore less likely to be true) than some alternative belief.

A different approach, which is more in line with the two strategies examined in this paper, is to claim that agents can sometimes permissibly fail to figure out the complete ranking. That is, agents are sometimes allowed to be in states of partial ignorance about the ranking. Figuring out the complete ranking would be rationally better, of course. But on the present view, it would be supererogatory.

Once we embrace the concept of epistemic supererogation, many of the theories we have examined become much more plausible. We might think, for example, that not being able come up with certain complex explanations is rationally permissible, but being able to come up with such hypotheses is much better. Thus, the beliefs resulting from the rationally better explanations are supererogatory. Or perhaps seeing that your evidence is inconsistent is not always required, given that the inconsistency is hard enough to see. But seeing the inconsistency is better – and hence, results in supererogatory beliefs. Alternatively, agents might be permitted to not always consider all of their evidence, although the epistemic saints who do are doing better, epistemically speaking. And finally, maybe ordinary epistemic agents only use one reliable epistemic standard, and this is okay. But if they considered all rational standards, they would be doing much better than okay – they would be epistemic heroes.

This is not to say that, on the final analysis, all of these views can be developed into a successful theory of supererogation. But every theory we have so far considered gives us a way into such a theory, if we are only willing to head in that direction.
8. Conclusion

We have seen that not only is epistemic supererogation a form of intrapersonal permissivism, it is perhaps our best hope for developing a plausible theory of this type. At least two promising strategies of developing intrapersonal permissivism turn out to suffer a common defect – the purportedly permissive situations they posit all require agents to be in a state of \textit{a priori} preventable ignorance. On a natural way of thinking about epistemic rationality, this is not a rational state to be in. The best way to patch up this defect is by creating conceptual space for a less than maximally rational, but still rationally permissible, doxastic state. In short, it requires the existence of epistemic supererogation. So our best hope for a theory of intrapersonal permissivism rests in a theory of epistemic supererogation. Combine this with the independent plausibility of each view considered in isolation, and with the mutual support the views lend each other when considered together, we end up with a package that has much to recommend it.

9. Works Cited


\textit{Philosophical Review} 122 (3): 337-393.


Elga, Adam, (ms). “Lucky to Be Rational.”


Titelbaum, Michael and Matthew Kopec, (ms.). “Plausible Permissivism.”


A THEORY OF EPISTEMIC SUPEREROGATION

Abstract: Though there is a wide and varied literature on ethical supererogation, there has been almost nothing written about its epistemic counterpart, despite an intuitive analogy between the two fields. This paper seeks to change this state of affairs. I will begin by showing that there are examples which intuitively feature epistemically supererogatory doxastic states. Next, I will present a positive theory of epistemic supererogation that can vindicate our intuitions in these examples, in an explanation that parallels a popular theory of ethical supererogation. Roughly, I will argue that a specific type of epistemic virtue – the ability to creatively think up plausible hypotheses given a body of evidence – is not required of epistemic agents. Thus, certain exercises of this virtue can result in supererogatory doxastic states. In presenting this theory, I will also show how thinking about epistemic supererogation can provide us a new way forward in the debate about the uniqueness thesis for epistemic rationality.

1. Introduction

In Arthur Conan Doyle’s short story “The Red Headed League,” Sherlock Holmes is faced with a strange case. A pawnbroker is accepted into a “Red Headed League” after responding to a newspaper advertisement brought to his attention by his assistant. The pawnbroker is asked to copy the encyclopedia for four hours a day, and is paid handsomely for his work. After several weeks of this, the league is suddenly closed and his supervisor leaves without a trace. Upon hearing this story and examining the pawnbroker’s residence, Holmes immediately solves the case. It turns out the pawnbroker’s assistant merely wanted to get his boss out of the house, so that he could tunnel into a nearby bank. His friend Dr. Watson, however, is completely befuddled:
“I trust that I am not more dense than my neighbours, but I was always oppressed with a sense of my own stupidity in my dealings with Sherlock Holmes. Here I had heard what he had heard, I had seen what he had seen, and yet from his words it was evident that he saw clearly not only what had happened but what was about to happen, while to me the whole business was still confused and grotesque.”

Let us assume that, as the story is written, Holmes and Watson share the same evidence – there is nothing relevant that Holmes knows which Watson doesn’t. Assume also that the shared evidence really does support the solution Holmes eventually came to. Holmes is a peculiarly brilliant epistemic agent – most of us would not solve the case along with Holmes, putting us in Dr. Watson’s shoes. Clearly this fact makes us epistemically worse than Holmes, at least in this particular situation. But does it make us “stupid,” as Dr. Watson says? Is our doxastic response to this evidence actually irrational? Are our beliefs about the situation actually epistemically unjustified?

There is a natural impulse to answer these questions in the negative. Maybe you feel stupid next to Holmes, we want to counsel Watson, but that doesn’t make you stupid! Holmes is a special type of agent who performs special epistemic acts – acts that involve levels of insight, intelligence, and imagination that even very rational agents can fail to achieve. But we aren’t required to exhibit such epistemic virtues.

Cases where agents exhibit extreme epistemic virtues are not exclusive to fiction. Take, for example, Albert Einstein’s theory of general relativity. This is considered one of the most important and surprising theories in the history of physics, and with its radical revisions of our view of the nature of space and time, it is not hard to see why. That Einstein not only seriously considered the possibility of such a theory, but was also able to show that it combines the predictive and explanatory power of many of its predecessors was an epistemic achievement of

---

54 Doyle (1927), 185
the highest order. But it seems that much of the evidence that supported Einstein’s theory was well known to physics of the time. Probably every sufficiently well-educated physicist was in position to justifiably believe in the theory of general relativity before it was actually discovered. If all this is correct, should we say, then, that all these scientists were actually irrational in failing to believe in the theory of general relativity? This seems far too harsh a verdict, not to mention insufficiently laudatory of Einstein’s great achievement.

It might be noticed that a very similar phenomenon arises in the realm of morality. Sometimes a moral saint or hero performs an act that most moral agents would refrain from — risking her life to save a stranger, for example. At first blush, there seem to be clear parallels between the moral case and the epistemic case. Like Watson next to Holmes, most moral agents feel morally inadequate when compared to our saintly brethren. The saintly acts often require moral virtues that most agents do not possess — extraordinary amounts of courage or strength of will, for example. But, we think, normal agents aren’t acting wrongly when we refrain from performing saintly actions — we aren’t required to exercise virtues to saintly degrees. Moral theorists, of course, have long written about these so-called “supererogatory” actions. Yet despite the relatively large and varied literature on moral supererogation, practically nothing has been written about its epistemic counterpart.

---

55 This case is famous among Bayesians as an illustration of the “problem of old evidence.” Specifically, Einstein argued that his gravitational field equations explained an anomaly regarding the perihelion of Mercury that was known at least 50 years earlier. This argument was instrumental in Einstein’s theory gaining traction among physicists in the late 1910s. Given all this, it seems that physicists had propositional justification to believe the gravitational field equations long before 1915. See Glymour (1980) for a discussion of how this case causes problems for Bayesians.

56 At the very least, it seems hard to avoid the verdict that scientists who were trying to explain the anomaly about the perihelion of Mercury were irrational. But even this seems too harsh a verdict.

57 Not all moral theories, of course, accept the existence of supererogation. Most notably, classical utilitarianism does not seem to have room for the concept.

58 An exception I am aware of is Hedberg (2014), although that paper is about epistemically supererogatory actions (such as gathering additional evidence or double checking past evidence), whereas I am interested in epistemically supererogatory doxastic states.
So our intuitions about particular cases give us reason to consider a theory of epistemic supererogation. Furthermore, there is also a theoretical reason to be interested in such a theory. There is a debate in the epistemology literature about the “uniqueness thesis.” This thesis states that for any body of evidence, only one doxastic state toward a given proposition is the rationally justified one. Many philosophers have found the uniqueness thesis to be intuitively implausible, but there are relatively few fully developed epistemic theories that can explain why it is false. Epistemic supererogation, if it exists, provides hope for another way of looking at the issue. Perhaps for certain bodies of evidence forming the ideal belief is actually not obligatory, even if it is “better” than forming any non-ideal, but permitted belief. Thus, there is a sense in which more than one doxastic state is a rational response to this body of evidence, and therefore, uniqueness is false. Thus, if we can develop a theory such that epistemic supererogation is plausible then we also have a new and promising way to deny the uniqueness thesis.

Finally, it seems that there are significant structural similarities between ethics and epistemology, the two most prominent normative disciplines within philosophy. To put it somewhat crudely, ethics is the study of what moral agents ought to do, and epistemology is the study of what epistemic agents ought to believe. Taking advantage of this seeming parallelism is a way to advance both fields. Thus, the fact that supererogation is an intuitively recognized facet of our moral lives at least suggests that we should look for a similar phenomenon in the epistemic realm. Of course, how far the analogy between ethics and epistemology goes is an

59 This formulation is from Feldman (2007), 148.
61 There are several papers arguing against the uniqueness thesis, but there are few positive proposals to explain how evidence can be permissive. See Ballantyne and Coffman (2011) and Kelly (2014). Perhaps the best extant view is the one defended in Schoenfield (2013). See Li (ms.) for a critical discussion of the Schoenfield-style view.
62 This idea is suggested, although not developed at length, by Douven (2009), 351-2.
63 This methodological point is also made in Berker (2013), 337-8, and Hedberg (2014), 3624-5.
open question. It is certainly possible that epistemic supererogation does not exist and a theory of epistemic supererogation is ultimately untenable. But even if this is the case, discovering this fact would itself be valuable, since discovering where ethics and epistemology come apart would illuminate aspects of both.

To these ends, in this paper I will develop a positive theory of epistemic supererogation that can vindicate our intuitions in cases such as that of Sherlock Holmes and Albert Einstein. Of course, there are possible alternatives to this theory of supererogation. One might even deny the existence of epistemic supererogation altogether, preferring to explain away our intuitions over vindicating them. We will briefly consider these possibilities at the end of this paper. However, my main goal is to show that a genuine theory of supererogation which respects our intuitions is an actual, plausible option.

2. The Good and the Right

In a 1958 paper, J.O. Urmson defined a class of actions that are “of moral worth” but “fall out of the notion of a duty and seem to go beyond it.” These are the actions later writers will call “supererogatory,” and much work has gone into a more exact definition of the concept. Since my interests are not in the concept of moral supererogation itself, I will stipulate a simple definition of supererogation. A morally supererogatory action, I propose, is an act that is (1) not morally required, (2) morally permissible, and (3) morally better than some alternative act that is morally permissible.

64 In this paper, I will avoid thinking about epistemic supererogation in a priori domains, if such a phenomenon exists. I suspect that the view developed here can be extended to such domains, perhaps with some modification, but the task is beyond the scope of this paper.
65 Urmson (1958), 205.
66 Plausibly, criterion (2) is redundant given criterion (3), since any act that is morally better than a permissible action is itself permissible. I include both criteria here just in case.
One way to understand this definition of moral supererogation is to think about the “two faces” of morality – what are often called the deontic and the axiological. The deontic face of morality deals with questions of duty and obligation. The axiological face of morality deals with questions of goodness and value. The concept of supererogation involves both these two categories. To say that an action is not required but permissible is to make a deontic judgment; whereas to say that it is morally better than some other action is to make an axiological judgment. To maintain the analogy, it seems the epistemic realm should also have two faces.

The axiological face of the epistemic realm is relatively easy to see. Doxastic states, it is often thought, enjoy different degrees of justification given a certain body of evidence. It is natural to think that the more justified a doxastic state is, the “better” it is. The deontic side, however, is more difficult. What we need is an understanding of concepts such as epistemic requirements and epistemic permissions. Accordingly, much of this paper will be an attempt to develop the deontic face in a plausible way.

3. Requirements and Epistemic Virtues

We can begin with what I take to be a paradigmatic case of ethical supererogation:

**Small Talk:** Dana is walking to work early one morning. Up ahead, she sees an acquaintance of hers, Fox, walking towards her. Dana doesn’t know Fox very well, but they have talked a few times in the past. She likes him well enough, although this morning she is completely indifferent as to whether she converses with him. As they get closer, Fox slows down as if he wants to make small talk. Cheerfully, he says “Good morning, Dana.” Dana considers quickly greeting Fox and being on her way. A quick greeting would certainly not be considered rude or even remarkable. But she is not in any particular hurry and it seems that Fox wants to talk. In order to be nice, Dana stops to engage in a few minutes of pleasantries.
In the story, it is clear that Dana is doing something that she is not morally obligated to do. Had she simply given a quick greeting and moved on, nobody would think that she had done something wrong. Fox is not someone that Dana has any special relationship with or commitments to, and failing to make small talk would not harm Fox to any appreciable degree. It also does not violate any socially agreed upon conventions. Fox would think nothing of it if Dana did not stop to talk. But when she did stop to talk, Dana also clearly did something morally better than not stopping. Fox’s demeanor indicated that he wanted to talk, and this was also the reason that Dana did indeed stop to talk – being nice is morally better than not being nice, all else being equal. Thus, I conclude that if supererogatory actions exist at all, it is clear Dana’s action is morally supererogatory.

This example is noticeably lacking in some of the drama of stock examples in the literature – often involving war heroes jumping on grenades. Those examples, however, are complicated by making the supererogatory action extremely demanding on the agent. While many cases of moral supererogation undoubtedly are of this nature, Small Talk shows that they need not be. And if we can proceed without this complication, we should. Of course, it might turn out that cases of demanding supererogatory actions are very different from cases like Small Talk.67 Even if this is true, cases like Small Talk still need explaining, and if the explanation can be carried over successfully to the epistemic realm, then we still have a theory of epistemic supererogation.

---

67 Portmore (2008) gives an argument to this effect – in cases where supererogation is generated by demandingness considerations, non-moral reasons override moral reasons. I don’t think that this explanation can be carried over to the epistemic realm. Roughly, this is because it does not seem like practical considerations can make any difference for epistemic rationality. The fact, for example, that forming a belief will have horrible consequences does not seem to provide any epistemic reasons to not form the belief. If this is right, then this is not a suitable model for developing a theory of epistemic supererogation.
To understand cases like **Small Talk**, writers such as Joshua Gert have proposed that there are two dimensions along which reasons, including moral reasons, can be measured.68 Adopting Gert’s terminology, the rough idea is that we can talk about which actions a reason requires and also which actions a reason justifies. These two dimensions can come apart in that the same reason might require and justify different actions. Dana’s moral reason for talking to Fox does not require her to perform the action. Indeed, we might think that it doesn’t require anything of Dana at all. The reason can, however, justify her gesture. Given that talking to Fox is also morally better than not talking to Fox, it seems that this case is a straightforward example of supererogation. Ethical theories with versions of this basic idea at their core have been defended by James Dreier and Michael Zimmerman.69 Generally, the idea is that supererogation happens only when the reason for the supererogatory action does not require that action, but does justify it. Thus, the agent is not required to perform the action, but is justified in doing so.70

While this might serve as a fine theory of moral supererogation, it is somewhat mysterious as it stands. What explains the difference between the two dimensions of these moral reasons? How is it that they can come apart in this way? To answer this question for the case of ethics, Dreier proposes that morality has two different “points of view” – we might think of them in terms of two different types of virtuous agents. We can call these the “just” (or maybe “dutiful”) agent and the “beneficent” agent. An action can be justified by reasons of beneficence, but can only be required from the point of view of justice. In Dana’s case, the

---

68 See Gert (2012), Gert (2007), Gert (2003). Though here I follow Gert in talking about reasons, it seems the same point can be put in terms of two different types of values (perhaps with some additional assumptions). See Zimmerman (1993) for an example of how this might work.
69 Zimmerman (1993), Dreier (2004). Zimmerman puts his theory in terms of two different types of values, though he admits this might generate different types of reasons as well.
70 It may be best to understand this theory as only a necessary condition for supererogation. This is because we need an additional criterion which states that the action also needs to be better than some permissible action (so that minimally permissible actions don’t get counted as supererogatory). Neither Dreier nor Zimmerman address this directly, and I will set the issue aside until later in the paper.
beneficence in making Fox slightly happier is what makes stopping to talk permissible at all. However, since making Fox happier will not make her any more just, that reason cannot require her to do it. The cases of supererogation that we are interested in, we might think, always work like this – they are always cases where justice cannot require the action, but beneficence can justify (and therefore permit) it.\footnote{It is unclear if beneficence can ever require any action. See Portmore (2008), 381, for reasons to think that sometimes it can.}

There is reason to think that this rough story can be transferred to the epistemic realm. Our two paradigm cases of epistemic supererogation have something in common. In both cases, the supererogatory belief was one that was incredibly surprising – so much so that it seems epistemic agents can be forgiven for overlooking their possibility. To even come up with the supererogatory belief as one worth considering is very difficult. After all, thinking up the remote possibilities that turned out to be the best explanation for the evidence is what made Holmes such an extraordinary epistemic agent. And what is most impressive about Einstein’s achievement was his coming up with such a radically different theory that fit all the data. These are the reasons the beliefs seemed supererogatory in the first place.

Note, however, that Holmes was not necessarily better at actually evaluating how well a given hypothesis is supported by the evidence. When Watson hears about the actual solution to the mystery, he finds that it really does make sense of the evidence – he can “put it together.” Similarly, even an undergraduate physics student can understand the theory of relativity and why it does indeed fit the evidence. Furthermore, if even after Watson hears about the correct solution he fails to understand why it explains the evidence, then Watson is being irrational. Similarly, a physics student who fails to see how the theory of relativity is supported by the data even after being suitably taught is being irrational. That is, after we are informed of the
supererogatory belief, it is no longer supererogatory. Intuitively, this is because the hard part –
coming up with hypothesis – has already been done for us.

This suggests two different types of epistemic virtues, to be used in an explanation
similar to Dreier’s. One is the more everyday virtue of seeing the support relationships between
certain hypotheses and a body of evidence. It involves figuring out how well a given hypothesis
explains the evidence, and ranking the available hypotheses according to their plausibility. This
is more of a housekeeping virtue, requiring something like analysis and critical reasoning. The
other is the virtue of coming up with the hypotheses themselves. This requires more creativity
and imagination.\textsuperscript{72,73} The rough proposal, then, is that epistemic reasons also come in two
dimensions. Doxastic states can be evaluated in regards to whether they exhibit the creative
virtue – this is analogous to an action’s being morally justified. Housekeeping considerations, on
the other hand, can require certain doxastic states.

It should be noted that in both these examples, the two virtues of creativity and
housekeeping do not work independently. Holmes would not be exhibiting the relevant sort of
creativity by thinking up crazy scenarios to explain his evidence. What makes a hypothesis
relevant seems to be a matter of how likely it is to be the belief ultimately justified by the

\textsuperscript{72} For a different discussion of this same distinction, see Nozick (1993), starting on p. 172.
\textsuperscript{73} Another way to illustrate this distinction is by thinking about Bayesian epistemology. The ideal Bayesian agent
has prior probabilities for all propositions. She does not need to do anything creative, instead merely taking in
evidence and updating on that evidence in a mathematically constrained way. And though she doesn’t have the
creative insights of Holmes, the ideal agent would still be able to solve the mystery of The Red Headed League.
Real agents, of course, do not have prior probabilities for all propositions. But it is natural to think that agents with
more priors, and priors that are plausible given their evidence, are epistemically better off than agents with fewer
priors. What made Holmes better than Watson was that he had the prior which turned out to be the solution to the
mystery, while Watson did not. Thus, the having of “good” priors can be seen as a model of the creative virtue,
whereas correctly updating according to Bayes’ rule can be seen as a model of the housekeeping virtue. Bayesians,
of course, tend to say little about an agent’s priors. This is a limitation of Bayesian approaches. Indeed, some of the
more famous problems for Bayesianism can be seen as failures to seriously contend with the creative aspect of
rationality. For example, the problem of old evidence can be seen as a problem with what happens, on the Bayesian
model, when an agent thinks up a new hypothesis that turns out to be well supported by evidence she already had.
See Glymour (1980).
evidence – how plausible it is in light of the evidence. But sorting out the plausibility of
different hypotheses is a matter of good housekeeping. Thus, the relevant sort of creativity must
in some sense be guided by good housekeeping. This interaction of the two virtues is something
that does not obviously occur in the ethical realm – so here, at least, the analogy comes apart.\textsuperscript{74}

Still, with this distinction in place, we have a way of thinking about epistemic
supererogation that is somewhat analogous to the theory in ethics. If such a theory also makes
sense of the intuitive cases, then we end up with a strong candidate for a theory of epistemic
supererogation. Before examining this further, however, we need to get clearer on exactly what
the creative virtue comes to.

4. “Coming Up” with Hypotheses

The creative virtue, we said, was a matter of coming up with the relevant hypotheses.
But what does this mean? What does it mean to say that Holmes “came up” with the solution to
the mystery while Watson did not? There seems to be a certain mental state, a propositional
attitude toward the hypothesis (which I will call P), that Holmes has but Watson lacks. The
creative virtue, it seems, involves coming to have this mental state. To get a better grip on this
state, consider the following variations of the Holmes case.\textsuperscript{75}

\textbf{Case 1}: Watson and Holmes are having a conversation before the pawnbroker comes in.
In conversation, Holmes tells Watson about P. Watson begins to think about whether to
believe P, but gets a phone call and it slips from his consciousness before he decides.
Five minutes later, however, the pawnbroker comes in and tells his tale. Watson is
immediately able to solve the mystery.

\textsuperscript{74} Another way these two virtues might interact, as suggested by Nozick (1993) p. 173-4, is when assessing the
merits of a particular hypothesis. Often, this will involve thinking up “its best incompatible alternative” (173).
\textsuperscript{75} These cases are inspired by an example from Friedman (2013) 170-1, which discusses the related (but distinct)
mental state of suspending judgment.
Case 2: As in Case 1, Holmes tells Watson about P but P slips from Watson’s consciousness without him ever forming a belief about P. In this case, however, the pawnbroker does not come for ten years. After the pawnbroker tells his tale, Watson is as befuddled as he was in the original story.

There is an intuitive difference between the two cases in regards to Watson’s mental state relative to the proposition P at the time of the pawnbroker’s entrance. This difference explains why Watson was able to solve the mystery in Case 1 but not in Case 2. I will say that in Case 1, Watson entertains P while in Case 2 Watson does not entertain P at the time of the pawnbroker’s arrival. Specifically, Watson entertains P throughout Case 1, beginning at the point that Holmes tells him about P. When the pawnbroker arrives, Watson is able to solve the mystery because he was entertaining the solution. In Case 2, however, Watson begins by entertaining P but stops entertaining it at some point before the pawnbroker’s arrival.

Note that, as I am using the term, P does not have to be occurrent in an agent’s conscious thoughts for the agent to be entertaining P. In neither case is Watson consciously thinking about P – a phone call occupies his thoughts before the pawnbroker comes in. Still, it seems that there is some sense in which the hypothesis is close enough to his thoughts such that he is able to solve the mystery in Case 1. In Case 2, on the other hand, the hypothesis fades away over the intervening ten years, such that Watson is no longer entertaining it by the time the pawnbroker comes. All this being said, consciously considering a hypothesis is certainly sufficient for entertaining that hypothesis. This is what is happening in the original Doyle story. Holmes considered the possibility of the Red Headed League being a way to get the pawnbroker out of the shop as he thought through the available evidence and how it fit together. That is how he came to entertain P, and this explains why he was able to solve the mystery.

Furthermore, entertaining is different from the classic doxastic states of belief, disbelief, and suspension of judgment. An agent need not have even considered the question of whether a
hypothesis is true for her to have entertained it. Entertaining is, however, compatible with all these states. In fact, I think that entertaining a hypothesis is something that is necessary for these states – as a precondition to taking any sort of doxastic attitude toward it. This would explain why in Case 2 Watson never formed any of these doxastic states, either before or after the pawnbroker arrived. Thus, we can think of entertaining P as a matter of having P “retrievable” or “available” for belief.

We should notice, though, that coming up with a hypothesis cannot just be entertaining a hypothesis that you previously had not entertained. As Cases 1 and 2 show, there are other ways to come to entertain a hypothesis. Though Watson came to entertain P, a hypothesis he had not previously entertained, I would think that Watson should not get credit for coming up with the hypothesis, since it was Holmes who told him about it. After all, coming up with a hypothesis is an exercise of the epistemic virtue, something that deserves epistemic praise. Clearly, what Watson did does not satisfy these criteria.

In these cases, it seems that Watson is not responsible for entertaining the hypothesis. That is why he does not get any epistemic credit for the feat. Thus, we come to an account of coming up with a hypothesis. An agent comes up with a hypothesis when (1) she comes to entertain a hypothesis she had not previously been entertaining and (2) she is responsible for coming to entertain the hypothesis. Clearly, for this account to be complete, we also need some conception of responsibility that applies to coming to entertain hypotheses. Developing such an account is beyond the scope of this paper, but hopefully most accounts of responsibility can be parlayed into the type of theory we need. After all, a plausible theory of epistemic responsibility should imply that agents only get epistemic credit for the beliefs that they are responsible for. So it is natural to think that whatever features of a doxastic state confers responsibility on an
individual also confers epistemic credit on that individual. But since agents seem to get epistemic credit for entertaining hypotheses as well, we will hopefully find the same credit-conferring processes posited by the theory in some cases of hypothesis entertainment. For present purposes, however, I hope that an intuitive understanding will suffice.

5. Requirements and Creativity

Let us take stock. Inspired by the case of ethics, we have two different types of considerations playing different roles in determining the epistemic status of a doxastic state. With this in place we can explain intuitive cases of supererogation. In our Holmes story, we can begin by explaining why Watson is not irrational in failing to come up with the solution. This is because Watson is not required to come up with the relevant hypothesis that constitutes the solution. The hypothesis is such a complex and unusual causal story that it would require a serious exercise of the creative virtue in order to think up. Given that Watson did not think up the hypothesis, it would not be good housekeeping to believe it. This does not mean that Watson is not bound by any requirements. He is required, for example, to refrain from believing that a lizard god was responsible for the Red Headed League. This is because such a belief would conflict with the housekeeping virtue. Finally, Watson might even be positively required to form some doxastic states. He might be required, for example, to believe that he has no idea who started the Red Headed League. On the other hand, Holmes did come up with the right hypothesis. This was not required of him, but he also did nothing epistemically wrong, since it was an exercise of the creative virtue. If we can then explain why Holmes’ belief was better than some alternative permitted belief, then we can explain why this is a case of epistemic supererogation.
However, to fully explain what is happening, we need to refine our understanding of what it is for a doxastic state to be permitted. After all, it should be noted that after Holmes has come up with the relevant hypothesis, seeing that it explains the evidence very well is simply a matter of housekeeping. But this means that if Holmes failed to form the right belief after already coming up with the hypothesis, then Holmes would be doing something which conflicts with the housekeeping virtue. This also seems to imply that Holmes is in fact required to form the correct belief in his situation.

At first blush, this seems like a difficult problem. After all, we claimed that supererogatory beliefs are not required. But there is really not a time at which Holmes’ belief was both possible and not required. Before he came up with the hypothesis, he was like Watson and failed to entertain it – since entertainment is a pre-condition for belief, Holmes was simply incapable of believing the solution. After Holmes came up with the hypothesis, however, he was required to believe it. So where, we may ask, did the supererogation go?

To answer this question, we need to make an important observation. Though we often think only of doxastic states as the objects of epistemic evaluation, there are clearly other candidates as well. Chief among them is the process of belief formation itself. And it seems that in Holmes’ case, the process that led to his doxastic state – a process that included coming up with the relevant hypothesis – was itself permitted but not required. This is because an essential part of the process is an especially impressive exercise of the creative virtue, insofar as Holmes is responsible for coming to entertain the relevant hypothesis.

This direct epistemic evaluation of belief forming processes can serve as the basis for an indirect way of evaluating doxastic states. In particular, we might say that Holmes’ belief in the right hypothesis was not required in the following sense: it was the product of an epistemic
process that itself was not required. Without knowing how an agent came to have a certain doxastic state, we cannot know whether the state itself was required in this derivative sense.

This sense of a non-required doxastic state may be derivative – it is dependent in the concept of a non-required epistemic process – but it is a perfectly coherent sense of the concept all the same. Indeed, we often say something analogous about doxastic justification. Roughly speaking, a belief is doxastically justified if it is both supported by the evidence (if it is propositionally justified) and formed “for the right reasons” in the “right way.” Someone who luckily forms a belief supported by her evidence through wishful thinking, for example, does not have a doxastically justified belief. In other words, a belief is propositionally justified but not doxastically justified when the epistemic process that resulted in the belief is defective. But this means that the idea of a doxastically unjustified belief is parasitic on the notion of a defective epistemic process. And there is no answer to the question of whether an agent is doxastically justified in believing P without also knowing how she came to this belief.

After the hypothesis has been entertained, of course, it can only be supererogatory if it also exhibits the housekeeping virtue. Thus, we can further refine our theory of supererogation. A doxastic attitude toward some proposition P is supererogatory when (1) it is the result of good housekeeping (2) it is the result of an epistemic process in which the agent is responsible for coming to entertain some hypothesis, and (3) it is epistemically better than some alternative merely permissible doxastic attitude toward P. Armed with our new sense of epistemic requirements, we see that any doxastic state which satisfies criterion (2) is also not required.
6. Too Much Supererogation?

Criterion (2) is satisfied when the epistemic process exhibits the creative virtue – but there is an apparent problem here. It seems implausible to think that, generally speaking, doxastic states which exhibit the creative virtue are never required. To take the most extreme case, suppose someone is so terribly lacking in the creative virtue that she cannot ever think up a single hypothesis. Even when presented with the visual evidence of a chair, she does not entertain the thought that there exists a chair in front of her. According to what I have said so far, this person would not be violating any of her epistemic obligations by not believing in the existence of the chair. In fact, the agent does not even need to believe that it seems as if there is a chair in front of her. The agent fulfills all her obligations as long as she refrains from forming any doxastic states. But this is absurd. Perhaps even worse, since supererogatory beliefs are those doxastic states resulting from a process that is not required, and there are no hypotheses that one is required to come up with, then any belief process that involves coming up with any hypotheses and satisfies good housekeeping results in supererogatory beliefs! So this suggests that, even in the actual world, supererogatory beliefs are everywhere – something that is clearly false.

To evaluate these worries, we can begin by remembering that to come up with a hypothesis, it is necessary that the agent is responsible for coming to entertain the hypothesis. Recall that Watson is not behaving creatively when Holmes simply tells him the relevant hypothesis. But this means that if the agent was not so responsible, then it is possible that the agent entertains a hypothesis without doing anything creative. In these cases, the agent could be required – via housekeeping considerations – to form certain beliefs without doing anything supererogatory. If it is plausible that this is what is happening in cases where agents are
intuitively required to believe certain propositions, then we have the explanation we are looking for: the beliefs are not supererogatory after all.

As a model for how this might work, let us look at the most troubling case: perceptual beliefs. When an agent has the experience of seeing a chair in front of her, it just is not plausible to think that she is not required to believe that there is a chair in front of her.\footnote{76 To find the most troubling case, imagine an agent who is dutifully attending to a chair in the middle of her visual field. Intuitively, she is required to believe that there is a chair in front of her. This is the type of experience I will be referring to throughout the discussion.} The problem, I think, is not just that it seems irrational to not entertain the relevant hypothesis, but that it is almost impossible not to. How can you have an experience as of a chair in front of you without the thought that a chair is in front of you even crossing your mind? This thought, I think, is on the right track.

The central idea is that sometimes, due to contingent facts about our psychology, we can undergo certain mental events that we are not responsible for. It is natural, for example, to think that we are not responsible for the kind of random thoughts that might flutter through our heads for seemingly no reason. The correct theory of responsibility, coupled with the correct description of our psychologies, will tell us exactly which things these are. If we allow ourselves to indulge in some psychological speculation, it is plausible to think that at least ordinary humans (with an average amount of experience with chairs) undergo a certain mental event that makes them automatically entertain the hypothesis that there is a chair in front of them upon seeing such a chair. Furthermore, whatever the correct theory of responsibility turns out to be, it seems likely that agents are not responsible for coming to entertain this hypothesis. After all, phenomenologically speaking, it doesn’t seem that we have to do anything before we come to entertain the correct hypothesis. Instead, the phenomenological character of the visual
experience just makes the entertaining of the hypothesis *come* to us. It thus seems like a non-agential event that happens to us, rather than an epistemic action that we can be responsible for. If this is right, then in these situations the entertaining of the hypothesis is not supererogatory. Thus, forming the relevant belief is also not supererogatory.

This strategy can be extended to other types of evidence. Testimony is an obvious example. Suppose you are a detective investigating a murder, and a witness tells you “Jones did it.” Upon gaining this evidence, it seems to me that the hypothesis “Jones did it” must be something you entertain automatically. Similarly, memory seems to be the type of thing that has a certain phenomenological character as well. So if memory is a distinct type of evidence, we might say that you entertain the hypothesis P automatically when you remember that P. If all this is right, then the problem of supererogation being too widespread in the actual world is avoided. In the actual world, all the agents we know about are subject to these kinds of mental events, such that they often come to entertain hypotheses in non-agential ways. And if these hypotheses are supported by their evidence, then good housekeeping requires the agents to believe them.

It is, of course, unclear exactly when an agent can count as epistemically responsible for coming up with a hypothesis, and there are many hard cases. What about, for example, if a hypothesis comes to you in a flash of insight, seemingly out of nowhere? Though we might be tempted to say that you don’t count as responsible in these cases, we also don’t want to say that any type of voluntary control is necessary for epistemic responsibility, since it seems unlikely we ever have this kind of control. Since I cannot delve into this complex issue in this paper, I must leave these questions unanswered.

It might be possible to make an even stronger claim. Borrowing from some theories in the philosophy of mind, we might think that certain phenomenological experiences have essential propositional content. See Burge (1986), Block (1990), and Shoemaker (1990) for versions of this thesis. If this is right, then we can say that an agent entertains a proposition simply in virtue of having an experience with that proposition as its content. Thus, it is not even metaphysically possible to both have that piece of evidence and not entertain its propositional content. Clearly, under these theories, entertaining these hypotheses is not something the agent is responsible for, and therefore, not supererogatory. Unfortunately, this theory has two drawbacks. First, it relies on a controversial thesis in the philosophy of mind. And second, it will not cover all the cases we intuitively want it to. Suppose, for example, that you see a wine stained carpet next to a wine glass laying on its side. Obviously, someone spilled the wine, and coming up with this hypothesis does not seem supererogatory. But clearly that proposition is not part of the content of the visual experience. So in these cases, we will have to fall back on the psychological thesis.
Of course, for all I have said, there are possible agents with different psychologies, such that they really would be responsible for coming to entertain hypotheses that we would consider extremely obvious. There are possible agents, for example, who would not automatically come up with the hypothesis that there is a chair in front of them when presented with the visual evidence of the chair. We might still have the intuition that even for these agents such beliefs should not be supererogatory. This intuition, however, can be resisted. For one thing, this question does not raise the “supererogation everywhere” problem – allowing that these merely possible agents have supererogatory beliefs in these cases does not imply that there is too much actual supererogation.

Second, it is unclear whether we can reliably transfer our everyday epistemic intuitions over to situations involving extremely unfamiliar agents. In general, it is not implausible to think that agents who are very different psychologically are also very different epistemically. Indeed, once we make the type of agent we are thinking of vivid, it seems that our initial intuitions are not so clear. Imagine a somewhat dim-witted species of alien who really does not automatically entertain the chair hypothesis. Imagine one of these aliens is especially creative, however. Upon seeing the chair, she reclines in her thinking chair with a pipe to slowly tackle the problem. After several hours, she finally realizes that there must be a chair in front of her. Her friend, Dr. Schwatson, is astounded by her mental acuity. Are we still so sure that this agent does not believe in a supererogatory manner? Or perhaps consider a more realistic example, involving agents with non-ordinary cognitive faculties. A thinker with autism, for example, might be unable to entertain hypotheses about social situations that many other people would entertain automatically. Intuitively, however, we might think that this agent is not required to form these

---

79 I owe this example to an anonymous reviewer.
kind of beliefs. And if she did manage to form such a belief, she might be believing in a supererogatory manner.

7. Epistemic Justification and Good Housekeeping

To recap, the theory we have arrived at is that a doxastic attitude toward P is epistemically supererogatory when (1) it is a result of good housekeeping, (2) it is the result of an epistemic process in which the agent is responsible for coming to entertain some hypothesis, and (3) it is epistemically better than some alternative merely permissible doxastic attitude toward P. To clarify this last criterion, I should say that by “alternative merely permissible doxastic attitude toward P,” I mean a doxastic attitude that the agent would be permitted to form had she not undergone the non-required epistemic process.

So far, we have spent most of our efforts clearing up criterion (2) and all that is involved in the creative virtue. In this section, we will explore the housekeeping virtue in more detail. Recall that good housekeeping is a matter of seeing which of the entertained hypotheses are well supported by the evidence, and forming beliefs accordingly. Some of the questions that arise about the housekeeping virtue – including questions of exactly what it means for evidence to support (or fail to support) a hypothesis – can be pawned off to a full theory of epistemic justification. But there is one question that is worth exploring here. Presumably, if no hypothesis is well supported, then good housekeeping would recommend suspending judgment on the issue. And if only one hypothesis is well supported, than good housekeeping would recommend believing that hypothesis. But what if more than one hypothesis is well supported? Is this even possible?
A natural way of thinking about the housekeeping virtue precludes the possibility of more than one well supported hypothesis for any body of evidence. Good housekeepers simply rank all of their entertained hypotheses in terms of plausibility given their evidence. They then only consider the most plausible of these hypotheses. If that hypothesis is plausible enough, then they believe it. If it is not, then they suspend judgment. There is no possibility of having this process output two different beliefs. First, the housekeeping virtue will clearly not advise embracing some hypothesis that is less likely to be true when there is a more plausible hypothesis being entertained. Second, two hypotheses can never be “tied” for the most plausible and both be well enough supported to be believed. This is because if two hypotheses are equally probable, good housekeeping will not recommend one over the other. Instead, it will advise suspending judgment between the two.

There are, of course, many ways to resist this natural line of thought. Maybe there is more than one way to rank entertained hypotheses by their plausibility that is consistent with good housekeeping. This difference may either be interpersonal (different for different agents) or intrapersonal (with the same agent having two different ways of ranking the hypotheses). Or maybe good housekeeping simply cuts agents some slack, such that they are not required to construct the entire ranking. This might allow them to be ignorant about the plausibility of some of their hypotheses. For example, agents may fail to notice that two different hypotheses are equally plausible while still performing her housekeeping duties well enough.

In many ways, this back and forth has been a rough recapitulation of the debate over the uniqueness thesis, but in a more limited form. The natural way of thinking about the housekeeping virtue that we started with is committed to what we might call “limited uniqueness,” which says that there is always only one uniquely rational doxastic state, given a
body of evidence and a set of entertained hypotheses. The alternative is to countenance the possibility that sometimes, with certain bodies of evidence and sets of entertained hypotheses, more than one doxastic state is sanctioned by good housekeeping. This might be because good housekeeping works on a threshold model, which allows believing any hypothesis that is “good enough.” For example, two different hypotheses might both serve as pretty good explanations for some body of evidence, even if one is a better explanation than the other. In this situation, perhaps a belief in either of the two hypotheses is good enough housekeeping, and agents are permitted in forming either doxastic state. Alternatively, it might be because that sometimes incompatible hypotheses can “tie” for best supported by some body of evidence. In these situations, maybe good housekeeping does not always advise suspension of judgment, but rather permits an agent to choose one of the two beliefs.

I personally find limited uniqueness very plausible, but I will not argue for it here. It is worth mentioning, however, that the model of good housekeeping which supports limited uniqueness captures many of the intuitions that drive philosophers into embracing the full uniqueness thesis. The “natural line of thought” that we started this section with is one picture that can motivate the uniqueness thesis. That is, if we think that rationality is a matter of ranking the plausibility of beliefs given a body of evidence and forming the most plausible belief, then uniqueness seems to be true. But if only limited uniqueness is true, then something like this process will be required anyway – under the guise of good housekeeping.

Thus, if limited uniqueness is true, we end up with a theory that can respect our intuitions for the stronger, general uniqueness thesis without fully embracing it. Given a set of entertained hypotheses, uniqueness is true, since the housekeeping virtue is so restrictive. The creative virtue, however, does not generate epistemic requirements. This allows for the stronger, general
form of uniqueness to be false, since there may be different sets of hypotheses that an agent can permissibly come to entertain, even given the same body of evidence. Thus, we can end up with the type of situations permissivists find so plausible – cases, for example, where two rational agents with the same evidence can end up with different beliefs. In short, we end up with a view that satisfies theoretical desiderata on both sides. This seems to be an attractive feature of our theory of epistemic supererogation.

8. Alternative Theories

We thus have a sketch of a complete theory of epistemic supererogation. In this section, we will briefly consider two alternatives to our theory. The first is a more straightforward way of thinking about supererogation, which does not require the complex theoretical machinery that we have spent most of the paper developing. The second is a way of rejecting the existence of epistemic supererogation by explaining away our intuitions in cases like that of Einstein and Holmes.

A simple way to develop an alternative theory of supererogation would involve positing a certain threshold of justification, above which agents are not required (but are permitted) to achieve.\textsuperscript{80,81} Thus, more than one doxastic state involving the same proposition could be above that threshold, even when some of those doxastic states would be better justified than others. Since any states above the threshold are epistemically permissible, agents would be permitted to form less than maximally justified doxastic states. Some doxastic states above this threshold—the ones that were better justified than other permissible states—would count as supererogatory.

\textsuperscript{80} Enoch (2010) mentions a view that may be developed in this way, see fn.9. \textsuperscript{81} I would like to thank an anonymous reviewer for pressing me on this point.
Such a theory, however, would incur a considerable explanatory debt. The theory posits a special point on a continuum of epistemic justification. In addition to more and less justified doxastic states, there is also a point which makes an axiological difference. Above this point, beliefs are rational to hold, while beliefs below this point are irrational. Such a threshold does not seem like it can stand as a brute fact – thus, more explanation is called for. Why is there such a threshold at all? And why is it where it is (rather than, for example, somewhere slightly higher or slightly lower)? It seems that to answer these questions in a non-arbitrary way, there must be some theoretical edifice which supports the existence of such a threshold.

Perhaps the best way to develop this theory is to look at the way we use epistemic words. Roughly, the idea is that in everyday contexts, when we use words such as “rational” or “justified,” we are not talking about maximal rationality. This is because ordinary humans rarely achieve anything close to maximal rationality – thus, it is an ideal that is irrelevant in non-theoretical contexts. Armed with this kind of epistemic contextualism, we can explain cases like that of Holmes and Watson. In ordinary contexts, Watson’s response is completely rational, even though Holmes’ response is better (and therefore, supererogatory).

This sort of view, however, tends to make a belief’s supererogatory status depend on relatively shallow features of our social or linguistic environment. Under this view, there is nothing epistemically important about the point of supererogation or this type of permission – that point is really more of a fact about how we talk about the agent, rather than a fact about the agent herself. In other words, it is a merely linguistic fact. It seems, however, that Watson really is doing something epistemically significant by discharging his epistemic obligations. Now

---

82 See Cohen (1987) for a theory that could be adapted in this way. Markovits (2012) presents a similar theory for ethical supererogation.
perhaps these epistemic intuitions can be explained away with mere linguistic facts, but failing to fully vindicate our intuitions is at least a cost of the theory. 83

In addition, a theory of this type would not represent a compelling form of permissivism. After all, everybody will admit that in everyday contexts we work with standards of rationality below maximal rationality. But this does not mean that there is anything theoretically interesting about these linguistic standards when theorizing about the structure of rationality. If this is all epistemic supererogation amounts to, then defenders of uniqueness will happily say that maximal rationality is what they were interested in all along.

So we see that there are reasons to be dissatisfied with this alternative theory of supererogation. Perhaps, then, the solution is to resist the urge to develop such a theory at all. One way to do this, while still taking our intuitions into account, is to claim that agents like Holmes and Einstein actually were rationally required to form the ideal beliefs in their respective situations. Ordinary agents like Watson, however, were not so required. This allows us to refrain from the judgment that Watson was irrational for failing to solve the mystery – a verdict that intuitively felt overly harsh – without talking about supererogation at all. 84

Why is it that agents like Holmes have different epistemic requirements than agents like Watson? The natural answer is simply that they are different types of agents. Earlier, we noted that it is plausible to think that thinkers who are psychologically different are also epistemically different. Perhaps the superior cognitive capabilities of epistemic saints like Holmes and Einstein make a difference for what is rationally required of them. Interestingly, an explanation

---

83 It is worth noting that analogous intuitions also exist in the case of ethical supererogation. When we think about supererogatory actions – such as donating all of one’s money to charity – we think that there is some morally important sense in which such actions are not required. We might think this also gives us reason to resist contextualist understandings of supererogation in the ethical realm. Indeed, in his original paper, Urmson distinguishes between doing one’s duty, even if the duty is so difficult that most agents would fail to do it, and performing a supererogatory action (pg. 200-1).

84 I would like to thank an anonymous reviewer for this suggestion.
of this type would also represent a type of permissivism about rationality. Because epistemic saints like Holmes and ordinary agents like Watson can have the same evidence but different epistemic requirements, uniqueness turns out to be false.

For this view to be a true alternative to the existence of supererogation, however, more needs to be said. To see this, consider what happens if Watson – against the odds – comes up with the solution to the mystery. It seems that he will have formed a belief that is rationally better than some other belief that was permissible for him to form. So it seems that the belief must be epistemically supererogatory for him. If this is right, then we need an explanation of why it is supererogatory, and why Watson was permitted to form the rationally worse belief even though a better belief was possible. In other words, we need a theory of epistemic supererogation.

Thus, to rule out the possibility of supererogation, we must say that it is not possible for someone like Watson to solve the mystery, rather than merely unlikely. We must say that Watson can solve the mystery if and only if it is required of him. This claim, however, gets us into theoretically fraught territory. It is a variant of the ought-implies-can principle – a principle that is fairly controversial within epistemology.\(^{85}\) Tying an agent’s epistemic requirements so closely with her cognitive abilities makes it almost impossible for agents with severe cognitive deficiencies to fail her epistemic obligations. This seems wrong because we often think an agent’s cognitive deficiencies explain her irrationality in the first place.

But even setting aside these theoretical worries, it just does not seem true that the only agents who permissibly form non-maximally rational beliefs are those agents who are incapable of forming such beliefs. Watson could have had a brilliant flash of insight out of the blue, for

---

\(^{85}\) See Feldman (2001) §3 for an argument against this principle in epistemology.
example. And those physicists down the hall from Einstein could have come up with the theory of relativity, even if it was unlikely. If some physicist of average intelligence just stumbled onto the answer out of sheer luck, we would not think that some psychological law was being violated. So this type of view, while avoiding the existence of epistemic supererogation, does not seem to correctly describe the cases.

9. Conclusion

To summarize, our theory of epistemic supererogation first distinguishes between two epistemic virtues – a housekeeping virtue of seeing the support relations between bodies of evidence and hypotheses, and a creative virtue of coming up with plausible hypotheses given some body of evidence. It then claims that a doxastic attitude toward P is epistemically supererogatory when (1) it is a result of good housekeeping, (2) it is the result of an epistemic process in which the agent is responsible for coming to entertain some hypothesis, and (3) it is epistemically better than some alternative merely permissible doxastic attitude toward P. This theory of supererogation vindicates our intuitive judgments in cases like that of Sherlock Holmes by finding a real place for supererogation in our epistemic landscape. Interestingly, our theory also parallels its ethical counterpart in many ways, while also providing a satisfying “middle way” between uniqueness and permissivism. All this is reason enough to think that developing a good theory of epistemic supererogation is a worthwhile quest.

10. Works Cited


*Philosophical Review* 122 (3): 337-393.


Li, Han (ms.). “The Trouble With Having Standards.”


