

A Theory of Epistemic Supererogation¹

Han Li

Pre-print Draft. Forthcoming in *Erkenntnis*.

The final publication is available at Springer via <http://dx.doi.org/10.1007/s10670-017-9893-3>.

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."²

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

¹ For comments on earlier drafts of this paper and discussions about these issues, I would like to thank two anonymous reviewers for *Erkenntnis*, David Christensen, Jamie Dreier, Tom Kelly, Christopher Meacham, Bradford Saad, Joshua Schechter, and participants of the 2014, 2015, and 2016 Dissertation Workshop at Brown University.

² Doyle (1927), 185

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 famous 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 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.³ If all this is correct, should we say, then, that all these scientists were actually

³ 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.

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.⁴

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.⁵ Yet despite the relatively large and varied literature on moral supererogation, practically nothing has been written about its epistemic counterpart.⁶

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.⁷ Many philosophers have found the uniqueness thesis to be intuitively implausible,⁸ but there are relatively few fully developed epistemic theories that can explain why

⁴ 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.

⁵ Not all moral theories, of course, accept the existence of supererogation. Most notably, classical utilitarianism does not seem to have room for the concept.

⁶ 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.

⁷ This formulation is from Feldman (2007), 148.

⁸ See Rosen (2001), 71, Schoenfield (2013), 3-7, and Kelly (2014) 298-300, for expressions of this intuition.

it is false.⁹ 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.¹⁰ 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.¹¹ 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 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.¹²

Of course, there are possible alternatives to this theory of supererogation. One might even deny

⁹ 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.

¹⁰ This idea is suggested, although not developed at length, by Douven (2009), 351-2.

¹¹ This methodological point is also made in Berker (2013), 337-8, and Hedberg (2014), 3624-5.

¹² 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.

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.”¹³ 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.¹⁴

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,

¹³ Urmson (1958), 205.

¹⁴ 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.

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**.¹⁵ 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.

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.¹⁶ 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.¹⁷ Generally, the idea is that supererogation

¹⁵ 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.

¹⁶ 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.

¹⁷ 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.

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.¹⁸

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 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.¹⁹

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

¹⁸ 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.

¹⁹ It is unclear if beneficence can ever require any action. See Portmore (2008), 381, for reasons to think that sometimes it can.

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.^{20,21} The rough proposal, then, is that epistemic reasons also come in two

²⁰ For a different discussion of this same distinction, see Nozick (1993), starting on p. 172.

²¹ 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,

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 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.²²

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

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).

²² 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).

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.²³

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.

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

²³ These cases are inspired by an example from Friedman (2013) 170-1, which discusses the related (but distinct) mental state of suspending judgment.

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.²⁴ 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

²⁴ 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.

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.²⁵ 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.²⁶

²⁵ 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.

²⁶ 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

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.²⁷ A thinker with autism, for example, might be unable to entertain hypotheses about social situations that many other people would entertain

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.

²⁷ I owe this example to an anonymous reviewer.

automatically. Intuitively, however, we might think that this agent is not required to form these 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, then 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.^{28,29} 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.

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

²⁸ Enoch (2010) mentions a view that may be developed in this way, see fn.9.

²⁹ I would like to thank an anonymous reviewer for pressing me on this point.

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.³⁰ 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 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.³¹

³⁰ See Cohen (1987) for a theory that could be adapted in this way. Markovits (2012) presents a similar theory for ethical supererogation.

³¹ 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*

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.³²

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 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.

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).

³² I would like to thank an anonymous reviewer for this suggestion.

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.³³ 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 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

³³ See Feldman (2001) §3 for an argument against this principle in epistemology.

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.

Works Cited

- Ballantyne, Nathan and E.J. Coffman, (2011). "Uniqueness, Evidence, and Rationality." *Philosophers' Imprint* 11 (18): 1-13.
- Berker, Selim, (2013). "Epistemic Teleology and the Separateness of Propositions." *Philosophical Review* 122 (3): 337-393.
- Block, Ned, (1990). "Inverted Earth." *Philosophical Perspectives* 4: 53-79.
- Burge, Tyler, (1986). "Cartesian Error and the Objectivity of Perception." In Philip Pettit & John McDowell (eds.), *Subject, Thought, And Context*. (Oxford: Oxford University Press) 117–36.
- Cohen, Stewart, (1987). "Knowledge, Context, and Social Standards." *Synthese* 73 (1): 3 – 26.
- Doyle, Arthur (1927). "The Red Headed League." *The Complete Sherlock Holmes*. (New York: Doubleday) 176-189.
- Douven, Igor, (2009). "Uniqueness Revisited." *American Philosophical Quarterly* 46 (4): 347 – 361.
- Dreier, James, (2004). "Why Ethical Satisficing Makes Sense and Rational Satisficing Doesn't." In Michael Byron (ed.), *Satisficing and Maximizing*. (Cambridge: Cambridge University Press) 131-154.
- Feldman, Richard, (2001). "Voluntary Belief and Epistemic Evaluation." In Matthias Steup (ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Oxford University Press. pp. 77—92.
- Feldman, Richard, (2007). "Reasonable Religious Disagreement." in Louise Antony (ed.), *Philosophers Without Gods: Meditations on Atheism and the Secular*. (Oxford: Oxford University Press), 194-214.
- Friedman, Jane, (2013). "Suspended Judgment." *Philosophical Studies* 162 (2): 165-181.
- Gert, Joshua, (2003). "Requiring and Justifying: Two Dimensions of Normative Strength." *Erkenntnis* 59 (1): 5 – 36.
- Gert, Joshua, (2007). "Normative Strength and the Balance of Reasons." *Philosophical Review* 116 (4): 533-562.
- Gert, Joshua, (2012). "Moral Worth, Supererogation, and the Justifying/Requiring Distinction." *Philosophical Review* 121 (4): 611-618.
- Glymour, Clark, (1980). *Theory and Evidence*. Princeton University Press.
- Hedberg, Trevor, (2014). "Epistemic Supererogation and Its Implications." *Synthese* 191 (15): 3621-3637.

- Kelly, Thomas, (2014). "Evidence Can Be Permissive." In Matthias Steup & John Turri (eds.), *Contemporary Debates in Epistemology*. (Malden: Wiley-Blackwell), 298-311.
- Li, Han (ms.). "The Trouble With Having Standards."
- Markovits, Julie, (2012). "Saints, Heroes, Sages, and Villains." *Philosophical Studies* 158 (2): 289-311.
- Nozick, Robert, (1993). *The Nature of Rationality*. (Princeton: Princeton University Press.)
- Portmore, Douglas, (2008). "Are Moral Reasons Morally Overriding?" *Ethical Theory and Moral Practice* 11 (4): 369 – 388.
- Rosen, Gideon, (2001). "Nominalism, Naturalism, Epistemic Relativism." *Noûs* 35 (15): 69 – 91.
- Schoenfield, Miriam, (2013). "Permission to Believe: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief." *Noûs* 47 (1): 1-26.
- Shoemaker, Sydney, (1990). "Qualities and qualia: What's in the mind?" *Philosophy and Phenomenological Research* 50 (Supplement): 109-131.
- Urmson, J.O., (1958). "Saints and Heroes." In A. I. Melden (ed.), *Essays in Moral Philosophy*. (Seattle: University of Washington Press), 196-216.
- Zimmerman, Michael, (1983). "Supererogation and Doing the Best One Can." *American Philosophical Quarterly* 30 (4): 373 – 380.