THE SUBOPTIMALITY OF MORAL DEFERENCE

David Nejezchleba

Mount Holyoke College

Department of Philosophy

May 2022

ACKNOWLEDGMENTS

First and foremost, I would like to thank my thesis advisor, James Harold. Without your guidance, wisdom, and flexibility this project truly would not have been possible. Thank you for your willingness to (almost) never meet at our regularly scheduled meeting time, and for offering invaluable feedback and critique. It has been a formative experience working with you on this thesis.

I would also like to thank Katia Vavova and Sam Mitchell, for continually challenging me academically and guiding me through the very early stages of the senior thesis process. You have both (positively) shaped my experience in the Mount Holyoke philosophy department, and I am immensely grateful for your advice and insight.

Thank you to my friends and family for your unwavering support during this process. Your encouragement throughout this journey was deeply valued throughout its completion.

Table of Contents

I Acknowledgements ……………………………………………………………..…2

II Introduction …………………………..…………………………………..……. 4

1 Literature Review …………………………..……………………………….…….. 6

1.1 The Moral Calculus App ………………….……………………………………6

1.2 Enoch's Instrumentalist Defense of Moral Deference …………………………....7

1.3 The Problem of Opaque Evidence………………………….………………..…9

1.4 The Importance of Moral Understanding……………………….………………10

1.5 The Value of Moral Understanding According to Hills…………….…………….12

1.6 Howell's Virtue-Centered Problem with Deference…………….…………..…17

2 Discussion of Kant, Reason, and Virtue …….………………………………….…..… 21

2.1 Introduction ……………………………..…………………………………...21

2.2 Kantian, Moral Understanding, and the Good Will …………….……………….…..21

2.3 The Role of Moral Authorities ……………………..………………………..25

2.4 Potential Problems for a Kantian Account of Reason……………………………..32

2.5 Moral Understanding and the Question of Virtue …………………………….....37

2.6 Objections and Discussion…………………………………………………....39

3 Conclusion ………………………..…………………………………..……..… 44

4 Bibliography …………………………..…………………………………..……..…..47

Introduction

When I first began this project, I began by asking broader questions about how *moral* expertise is distinct from general, non-moral expertise. Generally, we rely on experts in their field to tell us information we would not be able to know on our own. And we do not usually consider our knowledge-gaps as shortcomings – we have no intuitive or rational issue with the division of expert-knowledge in this way. Indeed, in my thesis proposal for this project, I cited an example of an individual trusting their mechanic friend when their car breaks down. They might consider themselves lucky to have an expert with them, and relying on them is not something about which they have to deliberate. Chances are, when you bring your car in for inspection, you believe what the mechanic tells you. If your dentist tells you that you need a root canal, or your doctor tells you that you have a broken bone, you trust their judgment. When checking the weather app on your phone, you trust the meteorologists behind the science and the experts behind the app.

But there is something perhaps different about moral deference – and there are a few ways to approach looking at the puzzle of moral deference. One might ask more epistemic questions surrounding testimony, the concept of "moral experts," and their role in conveying moral knowledge to laymen. There are also questions regarding how we might identify these moral experts, or if they can exist at all.

If we are allowed to sidestep our moral responsibilities, what might this mean for our moral practices in real life? Moral deference goes *beyond* individual cases, like asking your friend if you ought to donate to charity or vote for a particular presidential candidate. There are broader implications that follow from the permissibility of moral deference: take self-driving cars as an example. The skills possessed by self-driving cars are nearly excellent, but the idea that they might be programmed to make 'moral' decisions regarding civilian casualties might sound alarming. Should there be a 'moral authority' they are tied to? Certainly, this prompts less relevant questions about technology and robots, but questions of moral deference will inevitably be involved in cases regarding technological decision making as well as medical matters.

This thesis largely assumes the existence of moral experts and leaves the more epistemic questions unexplored. I do not investigate *what* moral knowledge consists of, that is, if there are 'objective' moral truths, and if they can be communicated by testimony. This thesis *mainly* focused on questions of moral worth, where it comes from, and the degree to which moral deference is *justified* even if it diminishes the moral worth of our actions. In chapter two, I move into questions surrounding where the source of moral worth lies from the Kantian framework as well as the virtue-ethicist framework.

I begin by introducing the idea of the *moral calculus app* to illustrate a more intuitive problem of moral deference – that is, what are our initial reactions to a society built on deference? There seems to be

a problem inherent in moral deference, and throughout chapter one we explore *if* this is the case, or if the problem with moral deference is actually a problem with *opaque evidence*. The idea that moral deference is permissible (and, in some cases, obligatory) is put forth by David Enoch, my main target. He provides an instrumentalist defense of moral deference in his piece *A Defense of Moral Deference*. Certainly, there is instrumental value in moral understanding as a means of ensuring individuals act well, and Enoch argues that this is the *primary* value of moral understanding. Alison Hills argues that moral understanding is valuable for more than procuring moral ends in *Moral Testimony and Moral Epistemology* – and I explore her position and the degree to which she makes a successful case for the *inherent* value of moral understanding. I then briefly delve into an alternative critique of moral deference from Howell, who argues that deference damages our virtue, or may indicate a lack of virtue in his piece *Google Morals Virtue, and the Asymmetry of Deference*.

In chapter two, I argue that if one takes a deontological *or* virtue-ethicist position, deference will diminish the action's moral worth. Moral understanding, I argue, is the source of moral worth. For the deontologist, this is because the good will *requires* or necessitates moral understanding. When one acts from a sense of duty, they act with understanding. Since a sense of duty is required for a morally worthy action, *understanding* is thereby required to perform morally worthy actions. I then explore to what degree *understanding* is necessary for virtue and one's development of virtue. I use Stichter's *The Skill Model of Virtue* to present the idea that understanding *is* required for virtue, and that from a virtue ethicist position, moral deference is *not* permissible. The conclusion will elaborate on some questions brought up by the literature, and potential avenues for further conversation.

Literature Review

Section 1.1: The Moral Calculus App

Imagine that the degree to which you are moral or immoral can be objectively quantified. In the near future, scientists and ethicists have developed software that allows you to see if you are a good person or not on the basis of your actions. A 'moral calculus machine' provides you with a numerical value of your 'moral score' after you perform a certain action. An app called the Moral Calculus App (MCA) shows you your 'score' based on the 'goodness' of your action, and this score contributes to the score of your overall moral standing. For example, if you give 20 dollars to a homeless woman, this action might be 9.5/10, and raise your overall score from 8.4 to 8.6/10. This app has taken the world by storm, and people everywhere are doing the 'right thing' to increase their score.

Certainly, the world may seem like a much better place if people have an incentive to perform the right actions. Individuals are now going out of their way to act like good, morally upstanding agents, and there is a net positive effect with regard to the benefits people enjoy. But something might be missing from this picture. When asked why they acted the way they did, people shrugged. Their 'right' actions were merely a means to bring up their score. When asked what made their actions 'right' from an ethical standpoint, they failed to articulate reasons. They did not have a coherent set of moral views, and they could not properly evaluate the actions of others either.

Consider what might happen if the app crashes. Nobody gets a score anymore, and they have a difficult time distinguishing good actions from bad ones. They also cannot compare this aspect of their social standing to others, so they are less motivated to do the right thing. Without the app to guide them, people return to acting selfishly. Perhaps some of them try to get it right, but they cannot properly articulate the moral reasons behind their actions. They lack an understanding of what morally relevant features bear on a situation, and have no moral cognitive skills on which to base their actions. Intuitively, there is something fundamental missing from these agents if they must rely on an app to do the right thing.

What the agents lack in Moral Calculus App is *also* what agents lack when they defer to moral authorities, that being moral understanding and moral affects (or intuitive moral perceptions). Certainly, an agent who does not know if stealing is immoral might defer to a moral authority. The agent may refrain from stealing, but could not give reasons as to why stealing is wrong, and does not have the moral 'affects' or emotive insights that a virtuous individual might possess. The reasons for which they act, that being testimony, lack contact with the morally salient features of the relevant situations. The agents lack the understanding necessary to defend their own reasons for acting, and lack the understanding necessary to evaluate the reasons of others. Moral understanding and moral affects are both necessary for an agent to be moral, or perform morally worthy actions.

Those sympathetic toward moral deference might initially worry about the risk of wrongdoing. After all, if moral authorities *are* more likely to 'get it right,' then if an agent refuses to defer for the sake of another moral value, this *increases* the risk of wrongdoing. We have no Moral Calculus App, but what we do have is moral experts, and if we are committed to the idea that we ought to reduce the risk of wrongdoing, then we should defer to them. Some hold the view that those who are exposed to ethical questions or have more experience with ethical questions and concepts are better equipped to deal with them, so it makes sense that they should be considered experts. In *Moral Experts*, Peter Singer argues that moral deference is not only morally permissible, but morally required given that certain individuals, namely philosophers, have more experience and therefore expertise in dealing with ethical questions or moral quandaries. Their experience renders them better equipped to deal with moral problems, so they ought to be considered *established* experts and laymen ought to defer to them.

The idea that there are identifiable and reliable experts to whom we should defer is a strong claim. This approach may open the door to some worries regarding elitism, or what role these experts play. Others defending moral deference might not go as far as Singer, and instead say that experts are just those who are more likely to 'get it right.'

Section 1.2: Enoch's Instrumentalist Defense of Moral Deference

Perhaps it is the case that we have a *minimum risk* to avoid the possibility of wrongdoing to a reasonable extent. There is higher value in *doing right* than performing an action with moral understanding. This is the heart of David Enoch's argument in *A Defense of Moral Deference*, where he makes the argument that since doing right has intrinsic value, there is nothing suboptimal about (or, there is nothing 'missing' from) an agent who morally defers to another. We care about being moral because we care about doing the right thing. If we fail to do the right thing for the sake of a moral quality in an individual, we would be acting immorally. It is this minimal requirement, minimizing the risk of acting wrongly, on which Enoch insists; it would be violated if one chose to make the moral decision for himself (Enoch, 13).

One's practical reasoning must appropriately prioritize the morally protected interests of others, and if one makes a decision that compromises these interests, one acts immorally. For example, Enoch illustrates for us an issue where minimizing the risk of wrongdoing is central and urgent to the outcome of the situation. Violence is erupting in the Middle East, and you and your fellow citizens are voting on important decisions being made in that region. Your vote will impact if bombs are dropped, if troops are deployed, and your decisions will risk the lives of citizens. You have a tendency to be unsure and inaccurate on important matters, and weeks after you cast your vote, you feel regretful or embarrassed

about the choice you made. You frequently question the views you held not long ago. Your friend, Alon, seems to hold stable views and have insight that you *lack* when it comes to making decisions on these matters (Enoch, 2).

Alon is also reliably right. On matters where you disagreed, you find out later that his views were more reasonable than your own. You very infrequently find him to be wrong, and he seems to possess moral insight that you lack. You and Alon come to opposing views about a recent conflict in the war, and you must decide how to cast your vote with the knowledge that real people will be affected by your choice either way. It seems like since Alon has gotten it reliably right in the past, it might be wise to defer to him. If we adhere to the idea that there is a minimal requirement that we must minimize the risk of wrongdoing, since moral experts are more likely to 'get it right,' it is not just permissible to defer to them in certain cases, but required of us. If you refuse to defer to Alon even though you *know* he is more reliable than you, not only do you risk accepting a false moral judgment, but you may risk doing wrong to others (Enoch, 13).

The case is not unrealistic; indeed, the decisions we make will often affect the lives of others, be that at a local or global scale. The stakes are certainly higher in Enoch's scenario, and the implication is clear that what we ought to be focusing on is reducing the amount of harm or wrongdoing that may come from our choice. The exercise of our virtuous character is not salient here, nor is the value in our own moral understanding particularly pertinent. When we know the expert is more likely to be correct, there is a lower risk of wrongdoing. In the face of moral uncertainty, deferring is the only morally acceptable option because the expert is more likely to be correct, thus the risk of wrongdoing resulting from our actions is minimized. To choose not to defer for the sake of some other 'moral good' would be putting the cart before the horse, or a misjudgement between the relationship of our values.

From Enoch's perspective, there is nothing wrong with the Moral Calculus App (MCA), since every agent is fulfilling the minimum requirement to avoid the risk of wrongdoing. In the MCA in particular, since agents will otherwise do wrong without the app, Enoch might say they are required to keep using it. It would seem to us, though, that there must be other features of a worthy moral action or worthy moral agent, and not only that they exist (as Enoch would agree) but that they are *as valuable* to us as minimizing wrongdoing. There seems to be an intuitive problem with Enoch's approach, because it seems as though there is something missing from agents using the Moral Calculus App, and this thing that they are missing is something that we value.

If one refuses to defer in a case where they *know* that another individual is more reliable than them, not only do they risk accepting a false moral judgment, but they may risk doing wrong to others. It is this minimal requirement, minimizing the risk of acting wrongly, on which Enoch insists; it would be violated if one chose to make the moral decision for themself. It is therefore morally required that

someone defer in cases of moral questions where the agent believes another person is more likely to be right. That is, if another agent's choice will result in less harm to others, deference is "not just permissible but rather is morally required" (Enoch, 15). One's practical reasoning must appropriately prioritize the morally protected interests of others, and if one makes a decision that compromises these interests, one acts immorally.

Section 1.3: The Problem of Opaque Evidence

The problem of moral deference, Enoch argues, is actually a problem with opaque evidence, or evidence unavailable to us. What we perceive to be a problem inherent in deferring to moral authorities is really an issue with forming beliefs on non-transparent evidence. By this, Enoch means evidence for the truth of an argument that says nothing about what makes it true (Enoch, 9). For example, if you were to observe a correlation between when it will rain and your aching joints. Your inference is correct, and when your joints start to ache, it always rains soon after. This is forming a belief on opaque evidence. If you formed your belief on *transparent* evidence (as opposed to opaque evidence), it might involve you knowing that changes in the air pressure cause your joints to ache, and this change in air pressure is what foreshadows the rain.

Enoch outlines a *moral* scenario where you observe a correlation (in others) between a wrong action and an increased heart rate through the trolley problem, which is a thought experiment that is well-known within the philosophical community. Imagine there is an oncoming train, and two tracks or paths it is heading toward. On the path it is heading toward, there are five people working. On the path beside it, there is one person working. You cannot yell at the workers to tell them to get off, they cannot hear you. You are, however, close to a lever that will divert the trolley from the five person track to the one next to it, thus sparing four lives. Utilitarians will favor pulling the lever, while deontologists will favor doing nothing. Each group has their own moral rationale for either pulling or not pulling the lever, and moral 'experts' are well-acquainted with these arguments.

If I see you pulling the lever in the trolley problem and observe an increase in your heart rate, I may come to the conclusion that it is immoral to pull the level. We assume that I have inferred the correct conclusion, and you criticize others who pull the lever as well. The issue here is one of opaque evidence, and the same discomfort we experience in this instance is the same discomfort we experience in cases of moral deference. Similar to the rain scenario, if you formed your belief on the idea that pulling the lever is morally wrong because we ought to never treat people as means to an end, then your evidence directly pertains to the truth of your argument and *why* it is true (Enoch, 9).

What reasons might someone have for rejecting deference to a moral expert in this case? The values of moral autonomy and moral understanding he classifies as *purported values*, and he distinguishes them as values that will never independently provide us a reason to risk acting immorally. What we want to be able to do is respond to the "morally relevant features" in situations where this is important or relevant. The reason we value moral understanding is because it allows us to do just that; when we gain moral understanding, we can act morally. To respond to a situation in such a way that permits the risk of wrongdoing for the sake of moral understanding would defeat the purpose of understanding. Since moral understanding is crucial to the moral worth of an action, some have posited that moral understanding is a worthy pursuit both for instrumental and intrinsic reasons. The question as to if moral understanding *has* value is less relevant to Enoch's point, which is how we value moral understanding in relation to minimizing the risk of wrongdoing (Enoch).

Section 1.4: The Importance of Moral Understanding

Authors skeptical or opposed to moral deference conceptualize the importance of moral understanding in a different way. While Singer and Enoch seem to conceptualize moral understanding as a means to reducing the risk of harm, others have argued that understanding has intrinsic value on its own, or has value equal to reducing the risk of wrongdoing. In *Moral Testimony and Moral Epistemology,* Alison Hills (2009) outlines the value of moral understanding and how this understanding contributes to the moral worthiness of our actions. More specifically, she outlines why an agent's relationship with moral understanding can be damaged through the act of moral deference, and how moral understanding is important for our ability to make moral decisions in the future. Our sensitivities might be severed by deference, which can impact our ability to do morally worthy, virtuous actions (Hills).

She begins by exploring the idea that there are those who think moral testimony is much like nonmoral testimony; what reasons are there to *not* defer if both moral and nonmoral knowledge can be acquired through testimony? If we agree that there are moral experts among us, people more likely to 'get it right' than ourselves, more knowledgeable, and more reliable, then it makes sense to defer to these experts. In nonmoral cases, we defer to others on their judgment alone, and we do not evaluate the reasons for their judgements, for the sake of acquiring knowledge. For example, you might not understand meteorology or wind patterns, but you believe the weatherman when he states tomorrow will be a gusty day (Hills, 95).

For the same reasons, we might assume that moral deference is no different. If moral knowledge can be delivered through testimony, and experts are more likely to get it right, then we ought to defer to them. However, Hills makes an important distinction between moral knowledge and moral understanding,

and uses this cleavage between the two to argue that one may have good reason to not trust testimony and reject deference to moral experts, though the advice of such authorities is acceptable (Hills). Separating moral knowledge and understanding in an important way allows us to explore why there is something wrong with agents who use the Moral Calculus App, even if they 'act right.'

How are we to conceptualize understanding and what it entails? Hills defines moral understanding roughly as the following set of abilities:

a) to be able to comprehend an explanation of *why p* when it is presented by someone else, and the ability to explain it in their own words

b) given the information, *q*, they can draw the conclusion that *p* (or that *likely p*)

c) in a case where *q* and *p* are slightly different, *x* and *y*, they can draw the conclusion *y* from the information *x*

d) if they are given the conclusion *p* (or *y*, for that matter) they could provide the correct explanation, *q* (or x for that matter) (Hills, 119).

To illustrate what moral understanding does and does not consist of, Hills presents for us three cases:

a) Sarah, who is five and refrains from lying on the basis of her parents say so

b) Peter, who is fifteen and disagrees with those who say that helping others is a moral imperative. Instead, he believes in natural selection and that the strong should trample the weak

c) Mary, who believes lying is wrong because it is disrespectful to people makes them unhappy; lying for one's own personal benefit is always morally wrong, but for cases where lying may make an individual happy she holds a more nuanced opinion. Sometimes the moral permissibility of lying is circumstantial, and she can apply her beliefs to a variety of situations to try and make the right call (Hills, 98).

She assumes that Mary's beliefs are correct, while Peter's are wrong. Peter, although he has reasoned about his beliefs, cannot achieve moral understanding because *understanding* is factive; you cannot understand something that is untrue. There are also situations where your own understanding is not transparent to you. If you are faced with a difficult moral decision and deliberate correctly about the relevant factors, even if you believe you have missed an important detail or doubt your reasoning, you may still possess moral understanding (given that you came to the right conclusion through your

reasoning). Mary, unlike Sarah, has also developed the ability to make her own judgements and reason about her own claims (Hills, 98). Mary can provide reasons and her explanation as to why lying is wrong, Sarah cannot; they may both *know* lying is wrong, but only the former possesses understanding. Hills points out that this exemplifies the difference between knowing *p* is distinct from understanding why *p*, but there is still more to be explored regarding the difference between understanding why *p* and knowing why *p* (Hills, 100).

Let's say we have Mary, and Mary's friend Skips, both of whom have been exposed to the rationale behind moral vegetarianism. Mary's understanding of the moral reasons behind vegetarianism are as airtight as her understanding of why we ought to not lie; she can provide reasons as to why we should refrain from eating meat, she has drawn the correct conclusions with the information provided to her, and she can apply her reasoning to different but similar scenarios. Skips is in a different position, where he can provide the reasons as to why one ought to be a vegetarian, but he does not grasp them. He cannot apply his reasoning to different scenarios, and he cannot, from his own reasoning, draw the correct conclusion from the premises. Skips *lacks* appreciation for the *reasons* as to why eating meat is wrong.

Hills states that "moral understanding involves a grasp of the relation between a moral proposition and the reasons why it is true," and this is distinct from believing that they are the reasons why X is wrong, and distinct from *knowing* that they are the reasons why X is wrong (Hills, 101). One must have a *systematic grasp* of morality generally; if you have this grasp, if you have adequate moral reasoning abilities, you can procure for yourself *new* moral beliefs that are correct independently of the judgment or rationale of others. Your ability to come to the right conclusion for the right reasons stems from *your* appreciation of the reasons, not a matter of luck or rote memorization (Hills, 100).

Section 1.5: The Value of Moral Understanding According to Hills

1.) Reliably Doing Right

To reliably do the right, moral action, you must possess moral understanding or a systematic comprehension of the moral reasons for the action(s). Perhaps it *is* the case you have a moral expert by your side, or that your intuitions never lead you astray, but this is very unlikely; moral decisions are very complicated, and involve a complex understanding of the problem if the answer is to be found. When solving moral problems we may run into in our lives, there will be instances where we do not have the time to consult a moral expert, and we must have adequate moral reasoning skills or moral understanding if we are to get it right (Hills, 106).

2.) Justifying Yourself to Others

The practice of reason giving, or the exchange of reason between parties, is very important to us; we are motivated to seek justification that others could not reject to defend our own reasons. Hills states that the minimum justification for our reasons involves being able to say what we were doing, and why we were doing it. If we cannot do this much, our reliability is diminished, and we are limited in our reason-giving ability (Hills, 106).

Imagine a hurricane has just hit your town, and you see a group of nearby people injured from a fallen tree. You decide to go help them, and see your friend among them -- he is not seriously harmed, so you tend to those who are seriously injured first. Your friend sees this and objects that since you are his friend, you should help him before anyone else. You reply that there are others who need help more than he does, so you have to make sure they are OK. This is an example of reason-giving or reason exchanging, and we do this to be understood by others. If I could not provide you with reasons as to why I was helping others first because I did not understand why I did this, this would put me in a strange or awkward place. You would be unable to assure your friend that you are doing the right thing.

3.) Virtue

Someone is *virtuous* when they reliably do the right thing, that is, when they are disposed to choose, feel, and act morally. According to Hills, being virtuous involves a complex set of dispositions as well as both cognitive and non-cognitive aspects. Non-cognitively, one must have motivations to do right, and the correct judgment to act rightly from these motivations, and these motivations must respond to moral reasons.

When you defer, when you ask *someone else* what to do, you are certainly forming beliefs about what is morally required of you, and you would be correct about what would lead you to do the right thing, but you are not *virtuous* because your beliefs do not respond to moral reasons. One conceptualization of a "virtuous person" might be that their reasons stem from their (correct) deliberation on the matter rather than another's testimony, but I will further elaborate as to *why* in section 2.4 in 'the skill model of virtue.' If your selfishness, for example, leads you to perform the right action, then your reasons and motivations are not moral ones (even if your action leads to a 'moral' outcome). When one does more than *act* right, that is, when one's thoughts, feelings, and actions are shaped or structured by responsiveness to morality (moral reasons, morally relevant features), then one may be considered a moral authority. Appropriate

> moral orientation necessarily involves utilizing one's moral understanding to form one's moral beliefs; testimonial knowledge does not meet this standard because *you yourself* are not oriented toward moral reasons, you are relying on another to be oriented for you. This is why moral *understanding* is a necessary, crucial aspect of a virtuous character (Hills, 107).
>
> To illustrate this, let's say Mordecai has just been appointed governor of Washington, and is responsible for signing laws into action, vetoing bills, and other such duties. He feels incompetent with regard to moral matters, those being sensitive cases involving values like justice or fairness. He cannot gain clarity through his own reasoning about the issues. Mordecai has a mentor, Rigby, who has experience as a former governor and possesses exceptional decision making abilities. Given that he is reliably right, he helps Mordecai make decisions and provides him with guidelines; when asked to speak on the issues, Mordecai provides the rationale Rigby gave him. Mordecai is a bad governor because he does not respond to the relevant features at hand, that is, he does not respond to the aspects of justice, fairness, or equity that we would normally expect. He might have the right motivations, and Rigby may give him the right knowledge, but there is a disconnect between his beliefs and the relevant features of the situation. For Mordecai to become a *competent governor*, he would need moral understanding; he would be able to form his beliefs independently, base them on the moral reasons, and he could apply his reasoning to similar situations in which these moral factors are relevant, (Hills, 110).

Hills further demonstrates the importance of 'the right reasons' by providing us with the example of "the knowledgeable extremist" -- someone who believes that it is not morally wrong to kill people unless they belong to a certain religion. His religious authority tells him to not kill his friend, who he wants to kill. Wanting to do the right thing, he refrains from murdering his friend. This is a case of someone who a.) is doing the right action (refraining from murder) b.) on the basis of his urge to do a moral action, or refrain from wrongdoing and c.) his testimonial knowledge that the action is the right thing to do (Hills, 115).

The knowledgeable extremist acted on the testimonial basis of his religious authority because he believed him to be reliable, but he did not act on the belief (morally salient feature) that his friend was a person who does not deserve to die. He had the right motivations (to do what is morally required of him, or morally right), and he acted with the *knowledge* from his religious friend that it is wrong to kill, but his action did not properly respond to the right reasons, making his action *not* morally worthy (Hills, 115).

Is morally worthy action (or acting with the right sensitivities to the morally relevant features of the situation) self-indulgent? There are those who may be interested in the right actions, but there are other reasons we may be concerned with doing the right thing; that is, depending on our circumstances in the situation, we may care about *the agent* rather than solely her actions. For example, if you're dying of dehydration in a desert and a stranger offers you water, you won't care if his reasons for doing so are 'good.' He might be concerned with getting into heaven or saving his reputation, but as long as you get the water his reasons will not concern you (Hills, 117).

This, however, seems to be a reflection of your own misfortune. It is an incomplete picture insofar as others might care about your motivations and moral inclinations as well as if you *understand* why that is the right thing to do. When we morally evaluate an action, it is necessary to assess and understand the reasons for the agent's action. We can think about the Moral Calculus App in the same way; it certainly has the effect of bringing about positive outcomes or ethical results, but it does not illuminate or generate the moral characteristics we would like to see in agents, and it fails to enable individuals to perform morally worthy actions. If we *only* cared about 'doing right,' we should have no problems with the MCA. But its consequence would be moral outcomes at the expense of agents with a grasp of their moral reasons for acting. The result of this is actions empty of worth, and agents lacking in moral substance.

Any party discussing moral deference seems to agree that understanding leads an agent to do what is *reliably right*, that is, understanding allows us to independently and accurately reason. Those less sympathetic to moral deference seem to stress the importance of *justifying* our moral reasons to others, as well as the importance of an agent's moral character and if they perform morally worthy actions. The justification feature will be explored in light of Philip Nickel's *Moral Testimony and Its Authority*. Nickel's view concurs with Hill's insofar as both propose that the action must stem from the agent's understanding (or recognition of the morally salient features). One must act from their understanding of moral claims, and must have an understanding of the morally relevant claims of an action, and this is what is required of morality. When we rely on another's testimony, we fail to meet this requirement for a morally worthy action (Nickel).

Nickel goes deeper than Hills to distinguish between different forms of moral deference and target ones that fall under what he calls *justificatory dependence*. This would involve Demonstrative Dependence, for instance when Kyle states *p,* the moral importance or salience of fact *h* is brought to light for Eric, who was previously *aware* of it but had not, until that point, thought it to be relevant. It would also involve Substitutive Dependence, which might look like Kyle stating *p*, and this giving Eric a reason

to believe *p*, and acts a replacement for Eric's non-deferential, independent/individual justification for *p*. There is a problem, Nickel thinks, with being influenced in this way that prohibits one from acting in a morally worthy manner. We will discuss justificatory dependence in more detail in chapter two when we discuss the *role* of moral authorities (Nickel, 255).

To illustrate this, let's say that your friend Irene is about to lose her apartment because she cannot pay the rent. She comes to you and requests that you loan her some money until she can get a job and get herself together. You know that Irene has a drinking problem and has had significant trouble keeping a job in the past, and while you have the money to give her a loan, you doubt she will change her long-term behavior. If you refuse the money and she loses her apartment, you believe she might seriously have to come to terms with her own alcoholism. At a loss of what to do, you go to your older, wiser friend Don. He tells you to refuse the money so Irene is forced to confront her problems and change her lifestyle. After Irene is forcibly removed from her apartment, she asks why you didn't lend her the money (Nickel, 256).

There is something intuitively wrong with stating "I didn't lend you the money because my friend Don thought that would be the right thing to do." There is something wrong with your *justification* of your actions (Nickel, 256). This example can also be understood from Hills' perspective, because one cannot reason-give, or justify themselves to others, which is a part of why we value moral *understanding* (Hills). Think back to the Alon case. You decide to vote the same way as your Reliably Right Friend Alon, and voting this way involves sending drone strikes to a war-torn country. After you cast your vote and they send the strike, weeks later a young man approaches you on the street. He asks which way you voted for that particular decision, and you tell him. He states that his mother was killed in the drone strike, and asks why you voted the way you did. You would be hesitant to reply that you voted the way you did because your friend Alon tends to be right about these things, and he would be dissatisfied if you were to give that answer. This is because, to Nickel, if we are to be virtuous agents, we cannot merely *want* to do the right thing, or have the correct motivations, we must also be able to recognize the moral features relevant to the situation and respond to those features accordingly (Nickel).

Nickel proposes the Recognition Requirement, where an agent must understand a moral claim and understand the relevant moral features bearing on the situation. When one defers to a moral authority, often they will fail to understand why the action was right. There are a number of other situations where an agent may act 'right' but fail to act for the right reason (Nickel, 266). A child that was punished for stealing does not stop stealing in the future because they understand why it is wrong, rather, they wish to avoid punishment. When they enter adulthood, perhaps they gain an appreciation (on some level) for moral reasons, but do not adopt those reasons as *their* reasons for acting. This person has been conditioned not to steal, but their actions to refrain from stealing (or even return stolen goods) are not

morally worthy. What morality requires is that we act on the basis of the morally relevant or salient features of a situation, and must incorporate our "moral cognition" in our decision making (Nickel, 266).

The "psychological and motivational makeup" of an agent, then, must be well or rightly-oriented. The individual must have a grasp of the relevant moral claims or arguments, and therefore must have an abundant array of moral concepts they already understand. The degree to which one's beliefs are systematized involves a degree of nuance that would pull us from the relevant discussion. What matters is that one *has beliefs* in a cohesive moral framework; when reasoning about an issue, their individual beliefs do not contradict each other. Their moral beliefs and values generally integrate into a coordinated, harmonized substructure of understanding (Nickel, 259).

Hills and Nickel's arguments rest on the idea that moral understanding is *necessary* for acting in a morally worthy manner. Indeed, how could one *act* right if they do not understand the reasons for which they act? A good test score is worthless if the student received their answers from the professor and copied them down. There is a greater degree of value for a score that genuinely reflects the student's understanding.

Section 1.6: Howell's Virtue-Centered Problem with Deference

In an important respect, Robert Howell differs in his conceptualization of understanding's relationship to virtue, or morally worthy actions. In *Google Morals*, Howell takes a similar skeptical position to moral deference to Hills, but de-emphasizes the role of understanding and takes a virtue-centered approach instead. He argues that since the beliefs one acquires from deference are estranged from the agent's moral character, then their beliefs and actions are not reflective of their virtues, and something is missing from the agent. The reason we *admire* good agents or consider agents *to be virtuous* is because their beliefs and actions reflect their character, or their virtues. We do not admire them for simply performing the right actions, nor do we admire them for having an appropriate *understanding* for their reasons, though it is important (Howell).

There are a number of issues with deference that Howell articulates from a virtue-emphasized perspective. The actions produced from deference sidestep moral character; we want to see actions and beliefs stem from the agent's moral qualities, and this is not possible if the actions are not born from the virtues of the agent. Deference may also result in an agent possessing a virtuous belief and performing an action that might be considered virtuous, but they fail to have the virtue itself. This presents an issue if we believe agents ought to have a coherent set of moral beliefs from which they reason. Deference-derived beliefs will not consistently be integrated into "the rest of the agent's present beliefs, and can fail to provide the proper ground for new beliefs" (Howell, 403). An agent cannot go on to attain "higher

degrees of virtue." The development of virtue is therefore limited, as the agent cannot act consistently or reliably.

None of this seems to conflict with Hills' perspective of deference. Howell, however, illustrates cases for us where one may possess moral understanding but still be morally deficient. He concedes that there is strength in the argument, but believes that it ultimately fails to provide an adequate story as to why deference is suboptimal. Consider the case of Sam:

> Sam has an extensive background of moral matters, and he has an unusually thorough understanding of different ethical approaches. However, he has 'two minds' -- a deontological one, and a utilitarian one. He understands both systems perfectly. When it comes to an ethical dilemma where the two systems disagree, he asks google morals (an app which tells you right from wrong) if deontology or utilitarianism is correct. Google answers with one of them, and Sam makes his choice. Here, we have a case where he deferred but his lack of understanding was not the issue, but the asymmetry still remains. The problem seems to be not with a lack of understanding, but with a deficit in another quality, or one's moral character (Howell, 397).

Is understanding necessary if one is to perform morally worthy actions? From Hills and Nickel's perspective, this would appear to be the case. But is understanding necessary if one is to be virtuous? From Howell's approach, it seems as though we can separate understanding from virtue, and that while understanding may allow us to *become* virtuous or develop our virtuous character, it is not *necessary* for virtue. To understand *why* Howell may conceptualize it this way, consider a savant, or a student who intuitively knew the answers to very difficult math problems. Or, take the case of mathematical genius Daniel Tammet. He experiences synthesia, where numbers appear as colors or shapes, and the sum of two numbers appears as an object that provides him with the answer. "When I multiply numbers together, I see two shapes. The image starts to change and evolve, and a third shape emerges. That's the answer. It's mental imagery. It's like maths without having to think" (Johnson).

If one can do math without having to think and still provide us with the correct answer, this seems like a problem for the *understanding* and justification condition. Perhaps Tammet cannot provide us with a conventional mathematical explanation for how he got his answers. Individuals who can do significantly advanced maths in their minds might possess an ability not gained from learning or understanding the way *most* individuals do. Perhaps one could reliably get the right answer morally 'without having to think' if they had strong enough moral intuitions. Just as there are those who are mathematically smart enough to 'understand' math in rapid, unconventional ways, there may exist those who can *morally* intuit in ways that do not require understanding as we have defined it thus far.

Consider the case of Ursula. She has no philosophical training, and she cannot articulate the reasons behind the moral actions very well. What she does have is a 'reliable disposition' to do the right thing, and she feels the appropriate emotions (affects) when faced with a moral situation, and she *knows* what the right action is. She can also recommend the right thing to others. When it comes to explaining the reasoning behind her actions, she cannot say anything beyond "I know it was the right thing to do because I *feel as though* it was the right thing to do, and I am consistently right in these matters." Others with conventional moral understanding can explain why her action was right, but she cannot. Still, there is something more admirable about Ursula, and she has something that blind deferrers *lack* (Howell, 399).

It seems the case that for some, *a priori* knowledge like mathematics or morality is highly accessible, and understanding in the way we normally conceptualize it is bypassed. Understanding *is not necessary* to reliably get the right answer, one simply has the mental or neurological tools to *know* (or feel) the right answer. While we generally understand the virtuous person to be able to articulate their reasons, this is evidently not always the case. The cases of Tammet and Ursula illustrate where understanding and being reliably right depart from one another. Virtue, then, might be divorced from understanding, and while we do care about one's virtuous character, it does not *always* necessitate understanding in order to achieve it.

It is undeniable that this case illustrates the importance of virtue and that it *is* something we care about. However, Hills does not deny this, and we *generally* understand a virtuous person having a reliable disposition to do the right thing to be a *better agent* when they can articulate the reasons behind their action, and more generally, we expect that they have the disposition that they do because they understand the reasons behind their action and have a well-integrated, consistent moral judgements. From Hills' perspective, we would expect that Ursula *could* explain the reasons behind her actions if she *is* virtuous, and if she could not, we would not expect her to have a reliable disposition.

The fact that Ursula cannot explain the reasons behind her actions seems to illustrate a potential problem for those centering *virtue* above understanding. We generally value understanding as an important aspect of virtue as most people *lack* the intuitive genius that Ursula (or Tammet) possess. For the average individual, we expect that they can justify their actions with reasons, and that moral understanding allows them to *become* virtuous. If we conceptualize virtue as a skill to be gained, in this way we can pose the Problem of Improvement. When one cannot articulate the reasons for their actions, certainly they cannot justify themselves, but moreover they cannot build upon their *own* understanding to improve their moral character.

Consider another Ursula, we'll call her Ursula2, who is very similar to Ursula. They both have 'reliable dispositions,' and tend to *feel* in moral situations and recommend the right thing to others. Ursula2, however, sometimes makes the *wrong* decisions: around 75% of the time she is right, and 25%

of the time she is wrong. After an inconsistent action is pointed out to her, she cannot articulate the reason why she did what she did. She also cannot articulate the reasons behind her good actions, as we have explored, so it would seem she does not have moral reasons at all. If she wanted to improve her moral actions, or choose 'right' 100% of the time, she would have no basis for improvement because she does not understand her own reasons. The *reason why her action was right* does *not* make reference to her explanation of her moral reasons. If one is less than extremely reliable, then this is an issue for the agent from a practical perspective.

Furthermore, while cultivating virtue or virtuous tendencies is certainly something we value, it does not seem to be the cornerstone of what we value when assessing the moral worthiness of an agent's actions or character. When a mathematics student fails an exam, he lacks an understanding of the material. Moreover, he also lacks mathematical abilities seen in those who have practiced for a long time, or mathematically born geniuses. What we would *expect* from him, however, is not necessarily genius or significantly advanced mathematical ability, though it would be admirable if he achieved it. What we would expect is that he achieves an adequate level of understanding which allows him to correctly complete the types of problems the exam demands. Similarly, we do not expect that individuals will have an advanced moral intuition like Ursula, but that they possess understanding in ways important to navigating their moral landscape and the problems within it. Illustrating the picture in this way allows us to see that while virtue is something to strive for, and developing a virtuous character is something we care about, it is not what we expect from each moral agent aiming to do the right thing, aiming to get it right, or aiming to perform a *morally worthy action*.

Discussion of Kant, Reason, and Virtue

Section 2.1: Introduction

The essential question of this chapter is *why* moral understanding has inherent rather than merely instrumental value from multiple positions within different moral or ethical theories. We will first explore the Kantian idea of the good will, why it requires moral understanding, and why it is intrinsically valuable. We will also briefly explore the Kantian idea of *duties to the self* as a possible requirement for moral worth, and how this is relevant to the idea of moral understanding as valuable outside of producing desirable or 'correct' moral outcomes. One's ability to self-scrutinize is something we value independently, and can enable us to better understand the role moral authorities *ought* to have in morally

guiding individuals. Additionally, we will explore some possible reasons we might have to distrust the Kantian account of reason by criticizing his *elimination argument*, and why we might conceptualize reason as a naturalist rather than a traditional epistemologist. If we reject the Kantian account of reason, we might reject the moral theories that follow from it – and some possible alternatives to establishing moral understanding as inherently valuable might be the virtue ethicist position. This will conflict with Howell's conception of a virtuous character, and we will discuss the differences in how philosophers understand virtue and its relationship to the individual themself.

Section 2.2: Kantian, Moral Understanding, and the Good Will

One's 'good will' is good without qualification, argues Kant. What *makes a person good* is his will that allows him to *make decisions on the basis of moral law*. In short, this means one acts out of a sense of moral duty or obligation. This is unlike ordinary or conventional notions of "she has a good heart," and means that an individual only makes decisions that she thinks are morally worthy (Johnson, Robert). She lets moral law guide her behavior, deliberates on moral considerations, and understands them to be good reasons for her behavior. The value of other qualities, like courage or intelligence, can be diminished in value in particular circumstances. For example, if one were to use their intelligence to create an atomic bomb, or if one's courageousness led them to fight for an unjust cause (Johnson, Robert). The value of a good will, however, is *always* desirable without qualification. It is unconditionally good because it relies on nothing for its value. One's understanding of duty is formed with respect for moral law, or an understanding of moral principles. When one understands moral principles, this is good regardless of the context in which it is applied. Kant provides the example of the shopkeeper to illustrate this point (Kant, 8) I expand on his example to illustrate clearly when one is or is not acting out of duty.

Let's say there is a shopkeeper named Eric, a shopkeeper named Keneth, and a shopkeeper named Stanley. Eric charges his customers fair prices because he knows that if he does, customers will keep coming back to his store. He has earned the reputation of a 'fair and just' shopkeeper that keeps the customers interested in shopping there. Kant would say that Eric is *not* acting morally because he is not acting with the right *intentions*; that is, he is not acting from a sense of *duty* to do the *right* thing. If he could get away with charging higher prices while maintaining his reputation as 'fair,' then he would charge his customers more. He is not treating his customers as *ends,* he is treating them as *means*. Similar to Eric, shopkeeper Keneth also charges fair prices, though his reasons are a little different. He cares less about his reputation and more about *feeling* good from performing the right action. When Keneth does the right thing even though he is under *no* legal obligation to do the right thing, he feels good about himself and his character – this is his primary motivation for charging the 'fair' prices. Again, if Keneth lost this

feeling, he would no longer have ‘reasons’ to do the right thing. His action is contingent on his feeling, not an understanding of the moral law.

In the case of shopkeeper Stanley, however, he charges fair prices because he knows that it is his duty to treat other people fairly. His customers are rational beings, and as moral beings they deserve to be treated as ends in themselves.They are not a means to a reputation, or feeling good about oneself, or becoming wealthy – they are moral agents that must be treated as such. With an understanding of duty in mind, shopkeeper Stanley can perform the right action and with the right intention, supplying his action with *moral worth*.

Adhering to moral rules is what makes other things worth pursuing. Pleasure and intelligence are worth possessing if one can acquire and maintain them without giving up their moral principles. The value of a good will, then, is not its ability to “secure valuable ends,” because it would be valuable even if “it were “completely powerless to carry out its aims” (Kant, 6). We carry out our moral duty when our motive for doing the action is guided by moral reason, in the case of shopkeeper Stanley. Our will is what provides the motive for our actions. When you act from your emotions, or desire to attain a particular end, then you are acting out of *inclination,* in the case of shopkeeper Keneth. When someone acts out of inclination, their actions reflect an emotional response – their actions are not done out of reason, but rather feeling. Perhaps, for example, someone is *prone* to be generous. They are inclined, then, to give more money to charity. This is not done out of an understanding that they *ought* to give to the poor, that is, their action does not stem from a rationale.

Why is an understanding of duty inherently valuable, independent of the ends it achieves? We, as a species, uniquely possess reason, and this reason is shared by all rational creatures. The moral law that *follows* from this ‘practical reason’ obliges or compels everyone equally (Kant). Since it applies to each being, we must treat each individual with the understanding that they *possess* this capacity – that is, one *acts* in accordance with moral law, and one *treats others* as autonomous beings inherently bound by the same law. Morality is contingent on our ability to reason, that is, what rational beings *ought* to do. If moral law relies on our ability to reason, it must be reason itself which has inherent value. According to Kant, reason, or one’s sense of duty, is the sole source of moral worth. This, I would argue, is a non-instrumental reason to accept the inherent worth of moral understanding. If the good will is unconditionally good, then moral understanding is unconditionally good insofar as it allows us to respect moral law. If adherence to duty is good *under any condition*, then that which allows us to *understand* duty is also good under any condition.

There are other, instrumental reasons that one could explain to support the importance of moral understanding, which may provide a more whole or coherent account for the significance of moral understanding. The instrumental reason is that reason can ground us in what we ought to do in a more

powerful way than ‘feeling’ or inclination can. For Kant, inclination is that which motivates our ‘animal’ desires, it is the push that we feel because we are, in part, animals, and we have bodily needs and desires. To act on an inclination is to give in to that animalistic desire. We have an *inclination* to drink water when we are thirsty. Similarly, we might have an *inclination* to give money to the needy, because our brains are wired to be *empathetic*. This is different from *reason*, which makes us consider and universalize the maxim from which we act. The kind of freedom you exercise when you respond to an inclination is not genuine freedom – one exercises freedom when they choose from *reason*. If I reason that eating meat is wrong, regardless of how I *feel*, I have reason to abstain from eating meat. If I base my actions on inclination, it is unclear what grounds this action, and it could be said that I have less reason to perform the ‘right’ action based on inclination alone. What if my ‘feeling’ changes, and I am no longer motivated to be a vegetarian? If I had reasons, I might be able to ‘override’ or evaluate my change in feeling, but if I lack reasons, I am more likely to do the wrong thing.

Instrumental reasons aside, the issue of self-scrutiny is also pertinent in considering the inherent value of moral understanding through a Kantian framework. If one lacks self-knowledge, or self-knowledge of his own good will, he has failed in a Kantian sense because he cannot scrutinize if his intentions (or reasons!) are ‘good.’ Intentions, for Kant, provide the source of the moral worth of an action – if one has a good will, their intentions are formed out of a respect for the moral law. One must have intentions that reflect his sense of duty, or “ the rules regarding what a rational agent ought to do” (Jankowiak). Think back to the shopkeepers – Keneth in particular does not operate from moral *reasons* – he operates from his inclinations. He cannot truly reflect on the moral basis from which he operates because he is not acting from *moral understanding*.

Does one have the right intentions if they wish to respect the moral law, but are unaware as to the specific ‘rules’ they ought to ground themselves in? Perhaps they do not know how to weigh moral considerations, or correctly apply moral principles. For example, imagine that someone beginning to study philosophy is considering abortion. He thinks that the pregnant woman is an ‘end in herself,’ and to force her to bear the child would be to use her as an *end* to produce life and her autonomy as an individual would be violated. But, he also understands that the fetus is (more or less) alive. Given that people must be treated as *ends in themselves*, the child’s life cannot be terminated for the sake of the mother’s autonomy – even if we value autonomy, we must adhere to the *categorical imperative.* We must respect humanity by respecting each person, the fetus included. However, the fetus does not have the *agency* the mother does. Must one be an ‘agent’ for their life to be respected? In that case, are we justified in killing babies – or just those who rely on another person to survive?

The agent cannot accurately apply the moral concepts he knows and come to a ‘Kantian’ conclusion. On some level, we could say he *does* possess moral understanding if he can successfully

argue for the correct Kantian standpoint on this, even if he does not *know* it is the 'correct' answer. Even so, he cannot weigh the importance of particular factors, and as such he does not possess *full* understanding. He is still interested in respect for the moral law, and in many cases he is able to apply concepts and do the right thing. In this case, however, he requires help. So, there are certain cases where *help* is necessary if an agent is to grow and further their understanding. Which kinds of dependence are justified, and does such dependence *still* take away from the moral worth of one's actions?

This also relates to a case in section 1.6, where Sam has a large degree of background on different moral frameworks, and his 'two-minds' come into conflict with one another. Though he understands both systems, he cannot distinguish the 'true' or 'correct' response overall, so he relies on 'Google Morals' to make his choice. Howell argues that the problem here is not *understanding*, but a deficiency in one's moral character. I would argue, however, that this case might parallel the one above, though not entirely. In both cases, there is an issue of *weight*, or *weighing* the correct moral reasons, or correct moral conclusions to act well. In Sam's case, the problem *is still* with understanding, or an understanding of correctly weighing responses to act rightly. Sam might have a high degree of understanding, but if one cannot self-reflect, if one cannot properly *weigh* their conclusions, the problem is still with understanding itself.

Perhaps it would be good, now, to explore the degree to which moral authorities can or should influence our decisions. If they ought to – it is at the expense of moral worth? Does the influence of a moral authority *always* diminish an action's moral worth? Below, we will take a look at some ways of potentially understanding these questions by examining different forms of deference.

Section 2.3: The Role of Moral Authorities

In *Moral Testimony and Its Authority*, Philip Nickel outlines several forms of moral dependence with varying degrees of deferece:

A.) **Heuristic Dependence**: *P* is a moral claim that Eric had not considered before, and once his friend Kyle states *p*, Eric considers it and thinks over it, and now is in a "position to have a complete justification and motivational disposition for it" (Nickel, 255)

B.) **Demonstrative Dependence**: When Kyle states *p,* the moral importance or salience of fact *h* is brought to light for Eric, who was previously *aware* of it but had not, until that point, thought it to be relevant

C.) **Substitutive Dependence**: Kyle states *p*, and Kyle's utterance gives Eric a reason to believe *p*, and this is a replacement for Eric's non-deferential, independent/individual justification for *p*.

D.) **Motivational Dependence**: Kyle states *p*, and the statement of the claim provides Eric with the "motivational encouragement" needed so that he acts on *p's* basis.

Nickel is concerned with "justificatory dependence" which refers to B and C, and when people come to believe *p* in these cases, they are not positions where they can act in a morally good way (Nickel, 255). One might argue that it might be possible to position the role of moral authorities in relation to these types of dependence where understanding is not compromised, and thus the value of the action being performed is not compromised. In chapter 1, we briefly discussed Nickel's Recognition Requirement, which is the requirement that an agent act on his *recognition* of moral reasons, or that his action is guided rationally from these moral reasons (Nickel, 266). For the sake of the argument, let's say that for an action to have moral worth, moral considerations and moral understanding *must* play a significant role in forming one's moral positions. That is, let us grant moral understanding as a requirement for moral worth. Where does this leave us in determining the role of moral authorities in guiding us, but not subtracting from our action's moral worth?

Nickel believes that both Heuristic Dependence and Motivational Dependence are not concerning because the agent still acts with understanding. In the case of Demonstrative Dependence and Substitutive Dependence, the role of the authority here is different. This is because *justificatory dependence* is occurring, where an agent relies on a moral authority to believe a moral claim for which they cannot provide an adequate justification. Given that "we require more of agents than merely acting in accordance with a correct moral belief," we *must* place limits on this type of deference (Nickel, 255). So where does this leave us with moral authorities?

It should be briefly noted that in certain situations, like the Alon case, there is an element of urgent, *necessary action* in moral decision making. This is true of most scenarios; in determining if it is moral to eat meat, you have until your next meal to deliberate before you *must* take action. In the Alon case, not only were lives at stake, but we can imagine that the next vote on military action was not far from the present. Action, in many cases, is very clearly pressing – one might *require* moral testimony immediately to make the right decision. One might argue that in more pressing or urgent cases, moral deference is not only permissible but *necessary*. It might subtract from the action's moral worth, certainly, but it is much preferred to performing a wrong action. In less pressing matters where one has *time* for

deliberation, we might say that they have a *moral responsibility* to deliberate, reflect, and gain understanding through this process.

Hills provides an example of a moral authority that enables an individual to become an *active learner* through a more (but not entirely) independent acquisition of understanding. For example, a moral expert might provide an individual with the case of the famous violinist who needs life support and then ask if it would be morally permissible to detach themselves from him. They might then ask about their stance on abortion and ask if the cases can be paralleled, or ask if "the same factors are present" (Hills 115). This is quite different from merely trusting the conclusion of the moral expert, like in Alon's case. In Alon's case, he is *not* using or developing his moral understanding. In cases where one is forced to compare cases, or reflect on previously held judgements, or *apply* moral concepts to different cases, they both *use and* improve their moral understanding. The moral authority has played a clear role in the development of the agent's moral understanding,

Does the role of the moral authority take away from the moral worth of the agent's action? In certain cases, there is a clear sense of diminished value in an action – for example, Alon's case. Other cases are less clear. An agent might *understand* the reasons for vegetarianism, and she might be able to defend her reasons very well by explaining abstract concepts and drawing comparisons to *other* cases with similar moral factors, even though she gained this understanding *through* a moral authority.

What category might the active-learner dependence fit into in Nickel's context? It might be *heuristic* dependence, in which case it is not *justificatory* dependence. The active learner is provided one example and is prompted to think about *other* examples that moral concepts may apply to, further prompting them to consider their intuitions and values. In chapter one, one criteria for moral understanding was the ability to come to the right conclusions when provided with *other* scenarios – by prompting an agent to apply concepts to multiple cases, they thereby develop and exercise their moral understanding. If this is the dynamic between the moral authority and the agent, then there is less of a problem with the action's moral worth being diminished or compromised. But might the active-learner dependence that Hills' describes potentially fit into *demonstrative dependence,* which *is* a case of justificatory dependence?

Nickel states that demonstrative dependence happens when one cannot discern certain moral facts, even though grasping these facts is necessary for making accurate moral decisions. For example, it is difficult to see when women are treated differently in conversation "topic or mode," and one who does not *understand* or see the *salience* of the facts of the matter cannot morally deliberate properly. When someone defers, however, or 'acts dependently,' the relevant reasons for their actions *are* in mind. One could compare this with two similar but distinct cases of Ignorant Ivan and Ignorant Igor. Ivan is unaware of the conditions of factory farming, but he recognizes that harming animals is bad. When the conditions

animals face under intensive-animal farming are explained to him, he understands that he should avoid buying meat from companies that partake in this practice. This is not justificatory dependence as there was no *moral* claim uttered, and the facts of the matter, once understood by Ivan, could be integrated and understood through his own moral framework.

But let's consider the case of Ignorant Igor, who cannot determine which practices 'count' as factory farming, or which farms are 'factory farms.' He is not aware of the *nonmoral* facts, that is, it is hard for him to see "the salience of these facts when enmeshed in various situations," and for him to rely on a moral authority for this information would be a type of *justificatory* dependence (Nickel, 264). It is clear that Igor will have trouble acting in a morally good way because his moral *understanding* was compromised in an important way.

Even if he forms the right intentions, if he has formed them from *deference* and cannot interrogate them, he cannot be his own judge. This is important to Kant, certainly, but it might also be important to non-Kantians, or those not strictly Kantians. The ability to self-interrogate is, for example, something we also value intellectually; when we are considering our own reasons for holding a certain political position, or reviewing our answers on a test, we value the ability to self-reflect on the reasons for doing what we did. We value reason, we value authenticity, and we value the ability to come to original and independent conclusions. This also applies to moral scenarios – if I am told that stealing is wrong, I can certainly refrain from the action but I cannot understand why. Hills thinks this is important in one's ability to produce reasons for *others*, or produce justifications for others as to why you hold that position. We could also say that without understanding, self-scrutiny is impossible.

One might object that even though we *value* these things, we only value them instrumentally, or insofar as they are conducive to long-run good decisions. I would reply that in the context of a *Kantian* framework, as we have been exploring, we can position these things as inherently or non-instrumentally valuable insofar as reason, or understanding, is valuable to our good will. Without intentions and a sense of duty, our actions do not have moral worth. Reason is not only unique to us, but it is suited to guide our moral thoughts and actions. Given that all rational beings possess the same faculties of reason and agency, we are all obligated to the same moral duties. Given that reason, in the Kantian sense, is suited for the development of one's good will, we are required to have a sense of duty if we are to be moral agents (Kant).

Important to the discussion is the distinction between perfect and imperfect duties as conceptualized by Kant. For Kant, perfect duties are ones that have no exceptions. In every occasion where it is relevant, you must carry it out. For example, the duty to tell the truth is a perfect duty. You have an obligation to tell the truth at every opportunity, because it *is* possible to perform this action maximally. It is possible to conceive of a world where this maxim becomes universal law *without*

contradiction, and where it is never detrimental to do it maximally. An imperfect duty is one where you are *not* required to do it every time because it would be *impossible* to do it maximally. The duty to help others is an example of this, because one could not help others at every opportunity without giving away *everything* or *every resource* they possess. This would mean that the individual would eventually lack the resources to help others. One is blameworthy when they violate perfect duties, one is praiseworthy when they fulfill imperfect duties (Johnson, Robert). Perfect duties are context independent, while imperfect duties are context dependent, that is, one must always follow perfect duties regardless of circumstance, but adherence to imperfect duties is dependent on one's circumstances.

Moral deference, in certain cases, could be thought of as an imperfect duty – we move toward understanding over the course of our life, but we might not *always* develop it at the expense of doing the right action. This view, I believe, is compatible with the idea of understanding's inherent value. Think back to Enoch's Alon case, where acting on one's own understanding or *limited understanding* may result in a less than optimal outcome, and one may be acting for the wrong reasons. Since they still care about doing the right thing, they defer, and this deference is justified, but the action is still suboptimal.

If you're doing math homework, and you require assistance to help you get the right answer, the way you went about solving the problem is less 'worthy' than someone who solved it on their own by truly and deeply understanding the material. Your goal is to provide the right answer to the problem – but merely being *told* the right answer is missing the point. In some non-moral cases, it's better to get the wrong answer and learn from your mistakes so you can achieve understanding. Indeed, there is no question that if your mentor helps you understand the problem, even though this is suboptimal, one's understanding is as important as getting the right answer itself. In moral cases, though, when harm is involved, it can be better to defer and 'get it right' because we have conceptualized understanding as an *imperfect* duty. It still contains inherent value, but this value cannot always be weighed against the value of preventing harm.

If, for example, you're doing math homework and you require assistance to help you get the right answer, there *are* cases where this action is justified if the development of one's understanding is aided by the mentor, and one would not have gotten the problem right otherwise. But, the way you went about solving the problem is less still 'worthy' than someone who solved it on their own by truly and deeply understanding the material. Even, in the moral case, if one agrees that there *is* intrinsic value in developing one's moral capacities but not at the expense of doing the wrong thing, this can still be 'bypassed' because we have still established that moral deference is *suboptimal* as long as *there is intrinsic value in developing one's moral capacities and understanding, independent from 'the right outcome.'*

We might be able to use the knowledgeable extremist as a concrete example as to the importance of *the duty to the self*, or the obligation to self-scrutinize. To do this, I would like to first introduce the idea of 'the moral Gettier case' before further elaborating on the relationship between self-scrutiny and understanding. The original Gettier case example is used to show that the definition of knowledge as strictly 'justified true belief' is flawed. A standard (non moral) Gettier case can be illustrated as follows: *the sheep in the field*: you are standing in a field, and you see an object that resembles a sheep in the distance. You believe that *in the field, there exists a sheep*, and this belief is true. But not because you saw the sheep – behind the tree there exists a sheep in the field, but the object you saw was a sheep statue. You possess a true belief, that there is a sheep in the field, and you are *justified* in believing this because you saw a sheep-like resemblance in the field (Hetherington).

The standard definition of knowledge is "justified, true belief" (Ichikawa). The Gettier case is one where the belief is true, and it is certainly justified, but it *falls short* of knowledge. Certainly, looking at a convincingly sheep-like object is justification enough to believe that there *is* a sheep in the field. Epistemologists generally agree that this falls short of knowledge, though, because something is missing (Ichikawa). I will not go further into *why* these Gettier problems fall short of knowledge because it is less relevant to the problem of moral deference. I do not have the space here to discuss different epistemological positions like internalism (justification is strictly internal) or externalism (justification is not solely produced by factors internal to the agent) with regard to this case, my point here was merely to illustrate an epistemological case comparable to a moral one (Ichikawa).

In a case of moral deference, an individual may desire to do the right thing, and through testimony he may know the right action. I am debating stealing medicine for my mother. I ask a moral authority, who tells me that stealing *is* justified in this context. So, I wish to do the right thing and reflect goodness in my actions, and I now know the right thing to do. My action still falls short of being morally worthy in a similar way to a justified true belief falling short of knowledge. Make no mistake, the parallel is *not* an epistemic one; the individual deferring cannot act 'for the right reasons,' so in a sense he only *happens* to act rightly. Similarly, there is an element of 'luck' in the Gettier cases (it just so happens that there is a sheep which makes the belief true).

Think back to Jeremy, the extremist member of radical religious-political X-D group. He believes that killing anyone outside of the group is morally obligatory. Jeremy knows Cindy, and believes he ought to kill her because she is not a member of his group. The religious leader of his group tells him it would be wrong to kill Cindy, and he decides to defer to his superior. Jeremy wishes to do the right thing, and he now *knows* killing Cindy would be wrong, so it "is not accidental, then, that he does the right thing"

(Hills, 115). In Jeremy's case, he does not act *well*, that is, he does not act 'for the relevant moral reasons' necessary to perform a right action. This is because he lacks understanding and a good will.

In light of the idea of the moral Gettier case, we can understand the knowledgeable extremist case from chapter 1. This case outlines an important distinction between having knowledge and motivation to do the right thing, versus having moral understanding of the right action. If one has the right motivations and *knowledge* of the right course of action, this looks a lot like a *moral* Gettier case – there is still something missing for this action to have genuine moral worth, just as there is something missing from the original moral Gettier case that causes it to fall short of knowledge.

It might be helpful to analyze the shortcoming of Kant's shopkeeper from this perspective. The shopkeeper is a little different in that he has a non-duty based motive, that is, he is motivated by his reputation. Unlike Jeremy, he may *or may not* be motivated to do the right thing. What these cases *both* illustrate is something important about the necessity of a good will if their actions are to have moral worth/moral value. What Jeremy and the shopkeeper *both* lack is an understanding of their sense of duty – Jeremy fails to know that killing is wrong for the morally relevant reasons, and the shopkeeper fails to keep fair prices for the sake of duty. A lack of moral understanding results in a lack of moral worth in their actions, just as the original Gettier case is not a case of genuine *knowledge*. Additionally, both Jeremy and the shopkeeper have failed in a Kantian way because they have *failed to self-scrutinize*. If they lack the self-knowledge of their own good will, they must ask a moral expert if their intentions are good. This means they are unable to be their own judge, meaning they cannot verify their own intentions, and have thereby failed at their imperfect duties.

In *In Defense of Moral Testimony*, author Paulina Sliwa argues that moral deference *can* and *does* lead to moral understanding. In fact, "testimony may sometimes be necessary in order to achieve moral understanding" (Sliwa, 27). Certainly, we can imagine cases where this both *is* or *is not* the case – an instance of a mother explaining why he ought to share is an instance where understanding is being conveyed, but in cases of pure moral deference, like the Alon case, there is no transmission of understanding. There is certainly something to be said for the role of moral experts in moral development, and one might consider them important in formative years when one cannot make moral decisions for themselves. The role of experts, or the *desired* role of experts becomes quite different in dealing with rational, autonomous beings – if we are approaching this from the Kantian framework, then one must be operating from a sense of duty, and if one is guided by a moral expert, then they are not guided by duty because the *reasons* on which they act are opaque to them. This is true in cases of *pure* moral deference.

In cases of deference where moral reasons are given or provided, there is a degree of nuance allowed depending on the type of deference utilized, as we have explored in light of Nickel's piece.

However, in cases of pure moral deference, or in cases where an individual is relying purely on testimony to make a decision – this is suboptimal, and even when an individual is 'training' their actions still possess less worth than an individual whose actions stem from their own recognition of moral responsibility. Furthermore, it seems Sliwa has failed to distinguish between knowledge from testimony as distinct from *understanding* from testimony, which can result in the knowledgeable extremist case. And as we have already explored, this is undesirable for the Kantian.

Hills argues that the virtuous person makes their own, independent moral decisions and judgements because she possesses her own moral compass and has a "rule" or set of rules. "She herself is responsive to moral reasons, in the sense that the explanation of why she chose that action must make reference to the reasons why it is right" (Hills, 111). The problem of deference seems to be that there is a disconnect between "second hand" moral knowledge and the *moral reasons* which make an action morally worthy or virtuous. When Sliwa states that an expert can inform you of the moral status of an action, that it's "wrong because it's cruel or unkind," then on the basis of that expert's testimony, one can then "believe both *that* your action is wrong and *why* it's wrong" (Sliwa, 26). The primary issue with Sliwa's argument is that she does not differentiate between knowing *p*, knowing why *p,* and understanding *p*. It is very possible (and likely) that one may come to know *p* and know *why p* on the basis of testimony, but testimony cannot deliver *responsiveness to moral reasons* that understanding requires. We can see Stichter illustrate this idea in *Skill of Virtue*, and we can see this idea expressed in Kant's idea of the good will. In Alison Hills' *Comment on Karen Jones and François Schroeter* (2012), she states that responsiveness to moral reasons necessarily involves "the exercise of your own ability to grasp why those features of your situation favor your action, and drawing the appropriate conclusion on that basis. In other words, acting for reasons requires you to understand why your action is `the thing to do' and to use that understanding in deciding what to do" (Hills, 234).

Section 2.4: Potential Problems for a Kantian Account of Reason

Kant states that reason is our "*practical* faculty," its purpose is to have an influence on our *will*. The will, for Kant, is our ability to align our actions with duty, or to conceive of our actions in a particular way. The "proper function" of reason is to produce our good will, which has inherent value. It is not merely good "as a means," for even if one was unable to *act* or fulfill the aims of their will, their will

would still possess value independent of our ability to use it. Similarly, a dollar bill still has value even if it is not being spent – it sits in my wallet, retaining its monetary value (Kant, 6). I am not so concerned with discussing Kant's conception of pure and practical reason, as I think these topics are largely irrelevant for the purposes of our discussion. I do, however, wish to discuss Kant's claims in Chapter 1 of the Groundwork with regard to his conceptualization of reason's "supreme purpose." We are grounding the inherent value of moral understanding in the Kantian idea of the good will, but we might have reason to *doubt* that 'reason' as the basis of a moral theory.

The following critique of Kant's view of reason is motivated by a naturalistic-epistemological approach, but the debate over the role of psychology in understanding reason (or, philosophical problems generally) long precedes this discussion. In *Neo-Kantianism and the Roots of Anti- Psychologism*, Anderson discusses the psychological wave that came across philosophical disciplines like "logic, epistemology, aesthetics, and ethics" in exploring what ways our mental capacities influence these domains (Anderson, 287). There are those, like the Neo-Kantians, who "insist that the psychologistic approach committed some fallacy," that is, they obfuscate *true* philosophical reason or 'knowledge.' As Anderson notes, Kant "advocates an anti-psychologistic conception of logic," stating that there are no empirical principles in 'pure logic,' so psychology cannot be *applied* to it (Anderson, 288).

Kant's 'good will' relies on reason, and this conclusion follows from his *elimination argument*, which I aim to discuss. His reasoning is as follows:

A) Nature has endowed us with faculties suited to the functions they are to perform

B) Instinct provides us with the means to survival and/or happiness; reason does not

C) If reason does not provide us with the means to survival or happiness, then it must have some other function

D) The function of reason must be what it is suited to perform

E) Reason is suited to the creation of one's good will

F) The will that is inherently good can be provided by reason alone (Kant, 7).

If nature has given us the faculties suited to the purpose, Kant then says since reason fails to provide happiness or means for survival, its purpose is *suited for the will*. He states that *if* reason were generated for us to "survive, thrive, and be happy," then nature would have *failed* in a significant way, and has created a "*very* poor arrangement" in designating *reason* for these purposes (Kant, 7). His claims, however, that instinct can better provide for us these things than reason, are unsubstantiated. Reason, understood through an evolutionary lens, or social epistemological perspective, serves a social function relevant to our survival and adaptation as a species (Mercier).

Philosopher Cosmides evaluates the role of reason in the context of human development. She argues that the "information processing mechanisms," that being reasoning, were not adapted for solving "arbitrary tasks" or suited for individual reasoning. Rather, they adapted as mechanisms suited for solving the particular "biological problems posed by the physical, ecological and social environments" in which our ancestors evolved (Cosmides, 188). Others have similarly argued that reason functions for the social purpose of convincing oneself and others. In understanding how one might interpret or decide the function of something, we might ask "what is it good for?" Darwin's theory provided a coherent foundation for an account of function as use – while inheritable traits of beings have multiple effects, the *overall* effects are typically more beneficial for reproductive success. When scientists discuss a trait's 'function' it is typically in reference to the 'beneficial effect' it has for an organism (Mercier, 178). This alone gives us reason to doubt Kant's conception of reason as 'pure logic,' separate from any or all empirical principles.

If we are defining reason as something for which it is useful, scholars agree that human reason is poorly oriented toward the "pursuit of knowledge and good decision" (Mercier, 180). If we follow Kant's logic, reason *would be* suited for survival, not the 'orientation of the will,' as he concludes. Kant demands that we utilize our *individual cognitive capacities* to establish our good will, or to self-scrutinize, or understand our duties to others. However, *this* does not follow given that reason is not suited for this purpose, as many social epistemologists would argue (Cosmides). Reason generally fails as "a general problem solving ability," or as "individual cognition" because the faculties of our minds were oriented to our early environments and the unique set of problems we evolved to overcome (Mercier).

This is best illustrated by the Wason Selection task, which illustrates that the problem is more easily solved in the context of understanding a social norm, or ensuring individuals are not 'cheating' the rule (Cosmides). When presented as a logic puzzle, individuals rarely provide the right answer.

The puzzle is presented as follows: You are presented with four cards, two with letters, and two with numbers. You are then presented with the following rule:

You are told these four cards have a letter on one side and a number on the other. You are given a rule about the four cards: ***If a card has a vowel on one side, then it has an even number on the other side****. You are asked, which card(s) do you need to turn over in order to determine if the rule is true or false?* (Bye)

The puzzle involves a bit of abstract thinking, or checking logical validity. In this experiment, individuals "*overwhelmingly* failed" to apply a logical method of falsification called *modus tollens* – they must look for *P* when *Q* is not present, because "the contrapositive requires there be no *P* if there is no *Q"* (Bye). Only cards showing either ~*Q* or *P* ought to be checked. Participants, however, failed to recognize this.

However, Wason's subjects did *quite well* if the experiment's 'presentation' or content was altered to determine if the rule was broken *in a social setting* (Cosmides). Subjects were presented with four cards, but this time they contained the words "beer" and "coke" and the numbers "19" and "35" (Bye). They were then presented with the following rule:

You are told these four cards represent patrons in a bar, and each card has their drink on one side and their age in years on the other. You are given a rule about the four patrons: ***If a patron is drinking a beer, then they must be 21 years or older****. You are asked, which card(s) do you need to turn over in order to determine if the rule is being followed?* (Bye).

Subjects, on this task, did overwhelmingly *well* on this task. They were able to reason using *modus tollens* in this scenario because the task requires rule-checking, or requires a social-exchange algorithm which humans have evolved to possess. This is because, as Cosmides notes, the evolution of our development was such that it "is advantageous to reason adaptively, instead of logically," that is, the human mind evolved to achieve frequently occurring adaptive challenges, and our mental-algorithms are domain-specific (Cosmides).

Our early environment demanded that we develop epistemic vigilance to spot liars and cheaters by developing a mechanism called *coherence checking*. Reason, as we typically understand it, is a "broad faculty" which allows us to solve a broad range of problems. However, social epistemologists have argued that our mind contains systems of "modular mechanisms" which allow us to solve particular problems (Mercier, 181). Cosmides puts forth the idea that we possess mechanisms to detect liars and cheaters, and

that “reasoning as classically understood doesn’t exist” (Mercier, 182). Others have taken a similar social-epistemological lens and put forth the idea that ‘reason’ was built for producing (and evaluating) justifications for others, which allows us to present our motivations and protect our reputations, as this was critical for our early survival (Mericer, 182). Reason, then, according to Mercier, was designed for producing justifications for the sake of social capital or mutual understanding, not as an individual tool designed to “guide our rational conduct” (Mericer, 183). Our species is a social one, and communication is central to our nature. Reason, then, can be understood as that which developed in *light* of our social nature, as a tool for cooperation, not necessarily for *moral* reasoning and understanding.

I do not wish to further evaluate the possible reasons for which reason could have evolved, or the mechanisms it may or may not consist of. We have clear evidence that reason may *not* exist as Kant conceptualized it, and this has larger implications for his theory surrounding morality, or *what* our moral plane of existence lies on. Kant conceives of reason as that which can produce the *sole* justification for our morality, and his notion of the good will relies on his conception of reason. He dismisses happiness as a potential ‘goal’ of reason, and in this he succeeds because reason *does* fail to produce happiness. However, it also fails in its logical domain, or purposes of individual cognition. If this is the case, if reason is not *suited* for enhanced decision making, then his conception of the good will can also be questioned.

Where does that leave us in the problem of moral deference? So far, we have been trying to establish that moral deference is suboptimal, because an action’s moral worth must stem from moral understanding of one’s actions and moral duties. If they defer, they do not possess moral understanding, and their action does not contain moral worth. Why, then, does moral understanding have *inherent* value as opposed to merely *instrumental* value? Kant’s conception of ‘the good will’ may provide an explanation as to why *understanding* is the source of moral worth. But, given the issues with Kant’s elimination argument, we run into some *potential* issues with his line of reasoning. If reason is not suited for purposes of individual cognition, then it does not follow that reason is suited for the good will. This could mean that we cannot *definitively* say that ‘the good will’ argument can be said to establish moral understanding as having inherent value. We will turn to other avenues to explore if moral understanding has inherent value in *other contexts*. That is, are there other ways that moral understanding can have inherent value, outside Kant’s conception of the good will? I think it can.

The motivation for this argument lies in the idea that moral deference hinders our moral understanding in many contexts. In cases of pure moral deference, it is the case that one’s ability to *understand* the significant moral elements at play is limited, and this understanding is the source of the

actions' moral worth. By deferring, they may have achieved the ‘moral’ outcome, but their action was *suboptimal*. Thus far, we have argued that moral understanding has inherent rather than merely instrumental value because of Kant’s conception of the good will, and the understanding that is required for knowing one’s duty. In the Kantian defense of moral understanding, it is clear that moral understanding is inextricable from one’s good will. However, we cannot fully rely on Kant’s moral philosophy because of the issues with *traditional epistemology* on which Kant bases his concept of reason. That is, Kant’s understanding of reason ignores the ways in which humans actually think and reason, and is mistaken regarding the purpose of reason itself. From a naturalist epistemological framework, reason is suited for social purposes in detecting ‘cheaters,’ or for reason justification.

The Kantian may not wish to give up so easily. Indeed, we might be able to say reason is *still* suited for its moral purposes, or that we can establish the value of the good will through other means. If we were Kantians, we may wish to explore these avenues, but this is not a *Kantian* thesis. For now, though, we can say that a Kantian approach may provide us with a framework from which we can establish the inherent value of moral understanding, and thus an explanation as to why deference is suboptimal, but we have reason to think that the Kantian approach is not *sound* after all. This is not to say that I wholeheartedly dismiss Kant, but I do think there is discussion to be had which may cast doubt on his system of thought. I believe if we can successfully argue that moral understanding has inherent value for other moral systems of thought, like virtue ethics, we can make a much stronger case for the inherent value of moral understanding. The reader may be convinced that regardless of the moral philosophy or framework from which one approaches the issue, moral understanding is something we care about enough to both value and foster for its own sake.

Section 2.5: Moral Understanding and the Question of Virtue

In searching for a potential non-Kantian reason for the importance of moral understanding, we might turn to the idea that *understanding* is necessary for virtue. It could be argued that *virtue*, or one’s ability to be virtuous, requires moral understanding. If it is the case that virtue itself has inherent value, outside of the ability to produce moral outcomes, then it would be the case that moral understanding also has inherent value.

Moral understanding may be necessary for virtue, or required in the process of developing one’s virtuous character. In his piece *Skill of Virtue,* Matt Stichter puts forward the idea that for one to become an expert, they must at first be a ‘beginner.’ To progress, they have to learn and progress to the next stage of development, where the process continues. This requires understanding, even if the *expert* is not acting on

the basis of 'reasons' and is instead acting by 'instinct.' If one is to make virtue 'second nature,' or intuitive, they must first possess understanding (Stichter). To achieve this level of expertise, one must first *learn*. Importantly, even if they do not 'reason' out their actions for each time they act, they still operate on the basis of their understanding. Hills' explanation of what is entailed by understanding involves an appreciation of *reasons* as to *why* an action is right or wrong, which is distinct from merely believing or knowing the reasons (Hills). Understanding necessitates appreciation, and an individual must grasp the reasons for a proposition in relation to the proposition itself. One must be able to independently explain the proposition and draw the conclusion *from* the relevant reasons. Not only that, but one possessing understanding must be able to draw a *similar* conclusion from a different scenario with similar important features. They must also be able to provide the right reasons based on the proposition (Hills).

It seems to be the case that attaining understanding in this way is necessary for virtue, or to move from beginner to advanced in achieving expertise. One cannot become an *expert* without first possessing understanding, or having *coherent reasons* as to why they are performing certain actions. Understanding, then, is *necessary* for virtue. One must begin as a novice, and they cannot bypass this stage of learning in order to become an expert. Pure moral deference prohibits learning and the development of one's character, and for this reason we should consider it *suboptimal*.

The reason understanding is integral to a virtuous character is because virtue, or intuitively feeling the right emotions about performing the right actions, is *second-nature*, and making these feelings, actions, and habits become second nature requires practice. This practice necessarily requires understanding, because doing the right thing *in response to the relevant moral reasons* requires understanding. In Chapter 1, we saw Howell (indirectly) argue that understanding is not necessary for virtue, and we will explore an alternative understanding of virtue and why it may *require* understanding. Certainly, one can exhibit a virtuous quality or action without being virtuous themselves – but to advance one's character, to *become* virtuous or practice virtue, one must possess understanding.

The stages of competence range from novice to expert, with the intermediate stages being 'beginner,' 'competent' and 'proficient' – stages we will briefly discuss. The novice begins to cultivate a skill by first engaging with aspects of the skill which do not necessitate previous understanding or knowledge. One develops 'simple rules' to decide what to do. To advance to the 'advanced beginner' stage, the individual goes from relying on rules to instead use maxims, which "take into account the new features of situations" of which they are aware (Stichter, 42). This seems to signify an advancement in understanding, or the start of one's grasp of relevant moral reasons in relation to the proposition. After a period of time, there will be an 'overwhelming' amount of maxims, so "hierarchical decision-making

process is used" as a ranking process to organize for a plan or perspective (Stichter, 42). This is competent proficiency. To step into the 'proficiency stage,' the problems are no longer approached from an 'observer' perspective, which draws from rules. Instead, he is "struck by a certain plan or perspective," demonstrating an integration of understanding.

The movement to 'advanced' beginner to 'competent proficiency' demonstrates a significant increase in the range of one's abilities to come to the right conclusions independently by means of their own understanding. They will be able to more coherently explain, and not only draw the correct conclusion from the set of premises, but draw *similar* conclusions in different (but similar) scenarios. When one achieves expertise, they see "what needs to be done in a situation and see how to do it, without needing a decision about how to do it like the proficient person." This reflects a complete grasp of the relevant moral features and this mastery is accompanied by deep, integrated understanding of the moral elements being dealt with (Stichter). Expertise is impossible without understanding, or having *coherent reasons* as the basis for their actions. Virtue, then requires understanding. It is *necessary* for one to develop their moral framework, the foundation of moral intuition. In a very meaningful sense, moral understanding is necessary for moral worth in both a Kantian framework and a virtue-centered framework.

We may need to briefly explore how we are defining virtue for the sake of this discussion. The Aristotelian conceptualization of virtue is that of excellence in reasoning or character. For example, courage is a virtue when applied in the right circumstances, but if one fails to properly apply it, they become rash or cowardly. Howell, in his exploration of the source of moral worth, defines virtue as having a "reliable disposition to act and feel in certain ways" (Howell, 403). If one has a virtue, they are acting from their *character*. Regardless of how we are *defining* virtue in this case, it seems that there are potential avenues to establish moral understanding as something *required* for virtue, or necessary for the integration of one's values.

Section 2.6: Objections and Discussion

Enoch, as well as Aristotle, might reply that one cannot be virtuous while they are asleep. Virtue, he might say, requires activity, for it to have moral worth, we must perform *good acts* to be virtuous. If the point of virtue is to do good things, then it seems we have only established instrumental reasons for moral understanding, and moral deference is not suboptimal because there is no real *loss* of moral worth in one's actions. Even if we imagined someone having moral understanding and virtue, but never doing anything good, then these things would have no value and *produce* no value. There is value in performing

the right action, but if you can't do that, then understanding and virtue are worthless. One might also make the claim that deference might still allow one to still develop their virtue, or deference *could* be used for the development of virtue instead of moral understanding.

One counterargument to this might be that the agent is still *virtuous*, in some sense, even if 'the sleeping life' is insufficient for a 'good life.' This idea will be explored at the end of the section. For now, we have two arguments against us: if virtue is not used, it has no value. Deference also may not prohibit the acquisition of virtue. Both of these arguments center around the fact that if *we only value virtue insofar as it allows us to attain moral ends*, then we are in trouble. The latter argument will also require exploration with regard to what role moral authorities *ought* to play in moral development. If we can make the case that virtue retains its value even when it is 'unused,' like a dollar bill sitting in a wallet, then we can make a stronger case for the inherent value of moral understanding through virtue. And if we can make the case that deference *can* and often does prohibit understanding in important cases, then we can make the case that, in certain contexts, moral deference is still suboptimal.

According to Aristotle's account of happiness, happiness itself is the highest good. We achieve happiness by setting reason in accordance with virtue, because as humans we are meant to reason *well*, and virtue is required for performing excellence. Virtue is not simply good because it allows us to attain 'good ends', it allows us to have inner balance of the soul. This is why deference is bad here too, because by prohibiting our understanding, we cannot develop virtues, so we cannot set our soul in accordance with reason, which is good in itself (Hursthouse). So, moral understanding is still inherently valuable, and deference is still bad. Why is virtue inherently good, or something we value independently of ends? Indeed, the virtue ethicist places virtue at the center of the ethical theory, and virtue is understood as something more than that which allows us to achieve moral ends, or help us understand rules (Hursthouse). If, as many virtue ethicists do, we place *happiness* or 'balance of the soul' at the center of importance, or that which virtue allows us to achieve, then deference becomes less desirable even if it allows us to achieve the right ends.

I will approach this subject from an Aristotelian perspective regarding virtue, but we will briefly explore other virtue-ethicist perspectives about the purpose and nature of virtue. The "virtue of something," in Aristotle's conception of the idea, is what allows an object to perform its "essential" action – for example, the paper-cutter cuts paper (Hursthouse). What makes humans unique, then, is our ability to reason. Both Kant and Aristotle identified *reason* as that which makes human beings distinct from other beings, and by extension, that which defines our 'purpose.' Our virtue, then, is that which allows us to reason *well* – Aristotle thought that 'being in a particular state of character' is what allows us to do this.

Virtue is that which enables us to enter this state, by understanding the 'best course of action' or the correct actions at the correct times. For example, one who has the virtue of 'courage' would not shy away from a battle they could win, nor would they jump headfirst into a fight for which they were unmatched. Rather, they would assess the likelihood of victory before acting courageously and fighting *well*. If we are to live our lives 'in accordance with reason,' then we must live life virtuously.

What does this mean for us? It means that we desire virtue, or, virtue is *valuable beyond* that which leads to desirable moral outcomes. In thinking of the 'courage' example, one might be fighting for a noble cause, and care about 'winning the battle.' It seems his virtue is only valuable insofar as it allows him to achieve the 'right' or 'desired' ends. Is action, then, necessary for the value of virtue to be 'realized?' Or is it more like Kant's good will, which retains its value like an unspent dollar bill? We can imagine a case of a political prisoner who has 'mastered' the virtues, but is unable to make choices which reflect his 'state of being.' He cannot make the correct choices as to *when* he should be patient, or how much *courage* he ought to display in a particular situation because he will not have the chance to *be* in those situations. With nothing to establish his virtue, can he be said to be virtuous?

In the case of the political prisoner, it might be argued that there are few instances where one is prohibited *completely* from acting in a virtuous manner — even the prisoner has the opportunity to practice virtue through self-expression by being truthful, through conversation by being witty, by practicing patience in frustrating circumstances, or courage during daunting ones. But, let's assume that the agent who has mastered virtue is in a long, deep sleep, or trapped in a deprivation chamber without opportunity to act virtuously. Could they still possess virtue? Yes, I believe they could, though the *value* of their virtue might be diminished. It might be said that one that has achieved a virtuous state has achieved something akin to Kant's good will, and though not *exactly* the same, they still exist as a state or 'condition' within an individual, and are still good *unconditionally*. The prisoner unable to *act* virtuously still has a 'state of the soul' that a non-virtuous prisoner *lacks*. For example, the sleeping monk/saint will still be regarded for the level of virtue he was able to achieve while he was awake, and he will not be regarded as non-virtuous while he is in a state of deep rest. Certainly, he will not be able to exhibit this virtue through his actions, and one might argue this *diminishes* the value of his virtue. Even so, a degree of value seems to remain – and his previous virtuous actions, thoughts, or feelings grant him a degree of value in his current state.

Both Plato and Aristotle described virtue as a "stable or unchanging facet of the individual" (Alison, 1). For Aristotle, one has virtue if their *soul* is in a certain state, and action is *not* required for the soul to enter this state. Rather, actions might seem to *reflect* virtue in the soul, though action (once virtue

has been achieved) does not *create* a virtuous character, it is an "overall property prescribed to the individual who is virtuous" (Alison, 1). For Plato, 'knowledge of the good' is what virtue is constituted from – is it not that one *is not* virtuous because he does not find himself in a situation which requires it. Rather, if he *was* in that situation, *he would be* virtuous. For example, a mathematics professor might not find himself solving an equation at all times. If he was given a problem or teaching a class, however, he would be able to solve the problem because he possesses the knowledge. Similarly, the political prisoner might have the virtue – whether that be 'knowledge of the good' or 'a state of the soul' does not matter – he does not need to *act* to be virtuous. A consequence of virtue might be that we perform the right actions, or 'act rightly.' We may need to act to practice virtue, or improve our virtue as individuals, but virtue *as a concept* does not require action. We desire virtue, or we *value* virtue independently of ends – meaning that it is good because it necessarily entails happiness, *eudaimonia*, or 'balance of the soul' (Hursthouse). One cannot have the balance of the soul without virtue, and vice versa.

Virtue must be *acquired* through practice and self control; Aristotle writes that 'the road' to being virtuous is paved by "exhibiting moderation in nearly every aspect of life," or acting as though one *already had* virtue (Alison, 2). So, to *become* virtuous, action is necessary, but once one has achieved this 'state,' it exists within the individual. I believe that if one were not able to act, the value of virtue could still be retained. We value virtue *not only* because it leads to good actions – certainly, the honest man will not cheat on his wife, but a non-virtuous agent could still perform honesty *without* being virtuous, as we have explored. As to if one could *become* virtuous without acting is a different question – and I believe the novice must *act* to attain *virtue*.

If virtue is valuable beyond allowing an individual to 'do right,' then where does that leave us? It seems as though if we value virtue, or place virtue at the *center* of our moral framework, then moral deference is *not* permissible in the way Enoch thinks it is. One might say that moral *understanding*, which is *required* for virtue, is valuable in that it allows us to achieve virtue – but this balance of the soul is something that cannot be attained by deference. One cannot 'defer' about the 'balance of their soul,' that is something they must acquire through practice and a sense of *understanding*.

In *To Kill a Mockingbird,* public defender Atticus Finch knew the racism in his community would prohibit him from winning the case for Tom Robinson, a Black man accused of assaulting a white woman. There was nothing that could have swayed the jury members to vote for the innocence of a Black man, despite the compelling case Finch made and despite the exonerating evidence at hand. Nonetheless, he provides a strong and honest defense for him, rightfully displaying courage by *doing what courage required of him*, or by doing what was right though it may have damaged his public image or put him (and

his family) in danger from members of the community. Atticus explicitly states to Jem that courage is “when you're licked before you begin but you begin anyway and you see it through no matter what” (Lee, 78). Atticus had nothing to *gain* from defending Robinson, and may have suffered losses by taking up this task – but he did it nonetheless because it was *virtuous* to do so.

Consider cases in religious texts in which an individual practices virtue beyond the ends it allows him to attain. In The Book of Job, a man named Job, acknowledged by God for his life of piety and consistent virtuous behavior, lives a life of fortune and prosperity. Satan convinces God to take away Job’s house and family, arguing that Job was only *good* because his virtue resulted in blessings and fortune. Job, he stated, was only good because he attained something good from his piety. After God destroys Job’s house, kills his family, and takes his means of livelihood away, Job continues to praise God – he “fell to the ground in worship” and said “The Lord gave and the Lord has taken away; may the name of the Lord be praised” (Job 1:20-22). It did not matter what resulted from his praise; even when his good fortune was taken, he was still virtuous for the sake of virtue itself. This logic appeals to our most basic intuitions about ‘virtue’ or good qualities. When we love someone, for example, we love them for the sake of loving someone, because we love them – and love is a *good* thing. We extend kindness to others not because we expect something from them, but because we value humanity, or because one might wish to honor humanity through his fellow man. We value the kindness extended to others when the agent *does not* expect anything in return.

Conclusion

We began this by exploring if moral deference is significantly different than other forms of deference. That is, if there is something strange about moral deference, what makes it distinct? Why might we be more willing to listen to the weatherman, but hesitant to defer morally? In chapter one, we explored the idea of a Moral Calculus app which would allow you to assess the degree to which your action was moral or not. There was no 'understanding' requirement to bring up one's score, and when asked why they performed certain actions, they could not provide a response. From here, we launched into a puzzle regarding the importance, both instrumental and intrinsic, of moral understanding and how this fit into the puzzle of moral deference. If moral understanding has value independent of the results it brings about, then the action will contain less moral worth if one is deferring. If moral understanding only contains instrumental value insofar as it allows an agent to reach certain moral outcomes, then there is no issue with moral deference – the action would not be suboptimal.

Hills' account of the importance of moral understanding only got us so far in establishing its *inherent value* – what Hills demonstrated to us was its importance in fostering the agent's relationship to making moral decisions, but largely in the context of moral ends. Certainly, this is useful to our argument – if we can establish moral understanding's evaluative purpose and importance for the agent, this will be relevant in the context of other moral or ethical frameworks. Hills, though, *does* discuss reason-giving (or justifying oneself to others) to highlight the importance of moral understanding. We might have strengthened this defense by arguing from Sperber and Mercier's position on reason – that it developed *to* allow us to more easily 'give reasons.' The same conception of reason used to critique Kant might have strengthened Hills' position in this way, and might be a potential avenue of further discussion.

Where might the issue with moral deference stem from? Enoch explains that it is the issue of *opaque evidence*, or operating off of evidence we cannot see or understand. If we come to a conclusion, usually we like to know its truthmakers. If we come to the conclusion that one cannot pull the lever on the

trolley based on the *heart rate* of an individual who has pulled it, then we are operating off of opaque evidence. The problem is not *specifically* with moral deference, then, but with lack of transparent evidence. What might Kant say about the issue of opaque evidence? He might agree that it is wrong, and one should perform actions based on duty, but this still does not let moral deference off the hook – actions performed off opaque evidence are still less morally worthy than actions performed out of one's sense of duty, given that duty required understanding.

In chapter two, I applied different lenses to underline the importance of moral understanding and explored the degree to which they succeed in doing this. For the Kantian, moral deference is suboptimal because moral worth stems from duty, or one's good will. For the virtue-ethicist, this may depend on how one defines or conceptualizes virtue – the degree to which understanding is required for *virtue* could also be disputed or further discussed in light of this. Howell himself did *not* think that understanding was required for virtue – in fact, he drew a very clear distinction between *understanding* and *virtue* – the latter of which he defines as reliably feeling and acting in certain ways, or possessing a reliable disposition to 'get it right' (Howell). When one has virtue, one acts "in and from" character, for example, the generous person will have deep intuitions about *when* it is appropriate to give, and they will feel 'generous,' or feel *good* about their action. In chapter two, I argue that *understanding* is required for one to be virtuous using Stichner's *Skill Model of Virtue*. I do not delve deeper into the case of Daniel Tammet in chapter one – which would have been an interesting discussion in light of this. *Can* it be said that Tammet, the mathematical genius, possesses understanding if he does not reason (or develop his method of reasoning) the way we might typically expect of a very skilled individual?

Where does this leave us? Enoch's arguments regarding the intrinsic vs instrumental value of moral understanding have become increasingly entangled with other moral theories in light of this discussion. If we value moral understanding insofar as it allows us to achieve moral ends, as Enoch thinks it does, then there is a relatively straightforward answer with regard to moral deference: moral understanding can be bypassed for the sake of the right outcome. However, the value of moral understanding might be intrinsic rather than merely instrumental if we apply a Kantian spin to the argument. From a Kantian perspective, moral understanding *does* have inherent value because the good will requires it. Moral deference will always be suboptimal, and an action performed by deference will contain less value than an action performed with a sense of understanding. However, as we have explored in chapter two, it is an imperfect duty, meaning one is not *always* required to act upon this duty. In weighing the value of moral understanding, one might reasonably state that if we are at risk of violating the minimum obligation to avoid the risk of wrongdoing, moral deference is justified.

There were some further concepts and topics that may have been interesting to delve more deeply into had this been a bigger project. For example, the naturalist response to the Kantian conception of

reason deserves more attention – but was not entirely related to the topic of moral deference itself. Nonetheless, the naturalist-epistemologist response to Kant's elimination argument was too strong to ignore, and weakened the degree to which a duty-based defense of understanding succeeded. The Kantian response to the objection – either as a defense of 'traditional epistemology,' or by abandoning the elimination argument altogether and discussing a different defense of the good will – may be a potential route of further discussion.

Other concepts regarding virtue may generate further discussion – how are we to define 'inner balance of the soul,' and what degree of understanding is necessary to reach this? There might also be further conversation to be held regarding if one can be virtuous while asleep. Perhaps it is the case that one can be virtuous in this state, but they are unable to live a good life. Would Aristotle completely agree? Would Howell? These are subjects on which we only briefly touched, and exploring them further in light of this question of moral deference would make for rich dialogue.

Bibliography:

Alison, John. "Human Virtue in Plato and Aristotle." University of Pennsylvania, 2011.

Anderson, R Lanier. "Neo-Kantianism and the Roots of Anti-Psychologism." *British Journal for the History of Philosophy*, vol. 13, no. 2, 3 Feb. 2005, pp. 287–323., https://doi.org/10.1080/09608780500069319.

Bye, Jeffrey K. "Psychology Classics: Wason Selection Task (Part I)." *Psychology In Action*, Psychology In Action, 17 Oct. 2012, https://www.psychologyinaction.org/psychology-in-action-1/2012/10/07/classic-psychology-experiments-wason-selection-task-part-i.

Cosmides, Leda. "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task." *Cognition*, vol. 31, no. 3, 1989, pp. 187–276., https://doi.org/10.1016/0010-0277(89)90023-1.

Enoch, David. "A Defense of Moral Deference." *SSRN Electronic Journal*, 2014, https://doi.org/10.2139/ssrn.2601807.

Hills, Alison. "Moral Testimony and Moral Epistemology." *Ethics*, vol. 120, no. 1, 2009, pp. 94–127., https://doi.org/10.1086/648610.

Hills, Alison. "Comment on Karen Jones and François Schroeter." *Analyse Amp; Kritik*, vol. 34, no. 2, 2012, pp. 231–236., https://doi.org/10.1515/auk-2012-0205.

Howell, Robert J. "Google Morals, Virtue, and the Asymmetry of Deference." *Noûs*, vol. 48, no. 3, 2012, pp. 389–415., https://doi.org/10.1111/j.1468-0068.2012.00873.x.

Hursthouse, Rosalind and Glen Pettigrove, "Virtue Ethics", *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/ethics-virtue/>.

Hetherington, Stephen. “Gettier Problems .” *Internet Encyclopedia of Philosophy*, https://iep.utm.edu/gettier/.

Ichikawa, Jonathan Jenkins and Matthias Steup, "The Analysis of Knowledge", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/>.

Jankowiak, Tim. “Immanuel Kant.” *Internet Encyclopedia of Philosophy*, https://iep.utm.edu/kantview/.

Johnson, Richard. “A Genius Explains.” *The Guardian*, Guardian News and Media, 12 Feb. 2005, https://www.theguardian.com/theguardian/2005/feb/12/weekend7.weekend2.

Johnson, Robert and Adam Cureton, "Kant’s Moral Philosophy", *The Stanford Encyclopedia of Philosophy* (Spring 2022 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2022/entries/kant-moral/>.

Kant, Immanuel. “Groundwork for the Metaphysic of Morals .” *Early Modern Texts* , Jonathan Bennett , 2017, https://www.earlymoderntexts.com/assets/pdfs/kant1785.pdf.

Lee, Harper. *To Kill A Mockingbird*. VINTAGE CLASSICS, 2020.

Mercier, Hugo, and Dan Sperber. “Why Do Humans Reason? Arguments for an Argumentative Theory.” *Behavioral and Brain Sciences*, vol. 34, no. 2, 2011, pp. 57–74., https://doi.org/10.1017/s0140525x10000968.

Nickel, Philip (2001). “Moral Testimony and its Authority.” *Ethical Theory and Moral Practice* 4 (3):253-266.

Singer, Peter. “Moral Experts.” *Analysis*, vol. 32, no. 4, 1972, pp. 115–117., https://doi.org/10.1093/analys/32.4.115.

Stichter, Matt. “The Skill Model of Virtue.” Philosophy in the Contemporary World 14 (2007): 39–49. Web.

Sliwa, Paulina. “In Defense of Moral Testimony.” *Philosophical Studies*, vol. 158, no. 2, 2012, pp. 175–195., https://doi.org/10.1007/s11098-012-9887-6.

Williams, Garrath, "Kant's Account of Reason", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/kant-reason/>.