SOME APPARENT OBSTACLES TO DEVELOPING A KANTIAN VIRTUE THEORY

AMY LARA
Kansas State University
alara@ksu.edu

Abstract

Several neo-Kantians have questioned the standard deontological interpretation of Kant's ethical theory. They have also responded to charges of rationalism and rigorism by emphasizing the role of virtues and emotions in Kant's view. However, none have defended a fully virtue theoretic interpretation of Kant's theory. I claim that virtue theory has much to offer Kantians, but that resistance to developing a Kantian virtue theory rests on faulty assumptions about virtue theory. In this paper I clear away three apparent obstacles to developing a Kantian virtue theory. The first regards his account of the virtues, which I argue is tangential to the issue of whether he can be interpreted as a virtue theorist. The second is Kant's codification of moral principles, which I argue is compatible with virtue theory. The third is the apparent explanatory primacy of the Categorical Imperative, which I argue is not fully supported by the textual evidence.

KEY WORDS: Virtue theory; Kantian ethics; John McDowell.

Resumen

Varios neokantianos han cuestionado la interpretación deontológica estándar de la teoría ética de Kant. También han respondido a los cargos de racionalismo y rigorismo, enfatizando el rol de las virtudes y de las emociones en la teoría de Kant. Sin embargo, ninguno ha defendido una interpretación plena de la teoría de Kant en términos de la teoría de la virtud. Yo sostengo que la teoría de la virtud tiene mucho que ofrecer a los kantianos, y que la resistencia para desarrollar una teoría de la virtud kantiana descansa en supuestos erróneos acerca de dicha teoría. En este trabajo, despejo tres aparentes obstáculos al desarrollo de la teoría de la virtud kantiana. El primero atañe a su versión de las virtudes, y argumento que es tangencial al hecho de si Kant mismo puede ser interpretado como un teórico de la virtud. El segundo es la codificación de Kant de los principios morales, respecto de la cual yo argumento que es compatible con la teoría de virtud. El tercero es la aparente primacía explicativa del Imperativo Categórico, respecto de la cual argumento que no está lo suficientemente sostenida por una evidencia textual.

PALABRAS CLAVES: Teoría de la virtud; Ética kantiana; John Mc Dowell.

In recent years many Kantian ethical theorists have questioned the standard deontological interpretation of Kant. Paul Guyer argues that Kant's theory "undercuts the traditional distinction" between deontological

and consequentialist theories (2000, p. 133). Allen Wood argues that Kant does not have a deontological view, if deontology is committed to moral rules that are not grounded in substantive ends (1999, p. 114). Barbara Herman (1993) has made the most sustained argument for rejecting a deontological reading of Kant in her essay, "Leaving Deontology Behind."

This re-examination of the structure of Kant's theory has coincided with an increased focus by Kantians on issues of virtue and character.[1] This new focus on virtue has largely been a response to recent criticisms of Kantian ethics from Aristotelian, Nietzschean, and feminist camps. These criticisms have been directed at the rationalistic, impartial character of Kantian ethics, and Kant's ethical theory has been charged with being alienating, psychologically naïve, and anti-emotional.[2]

In response, defenders of Kant have sought to show that the Kantian moral agent is emotionally connected to others, sympathetic in feeling, and not obsessed with following abstract principles. They have tried to show that Kant has a rich view of the psychology of the good moral agent, a view that includes proper emotional responses to situations, perceptiveness and sensitivity, and a well-developed character. In making this argument, Kantians have found it useful to undermine the traditional reading of Kant's theory as deontological, yet few have then tried to place Kant within the usual tripartite division of theories: consequentialist, virtue theoretic, or deontological.[3] Instead, they have explicitly or implicitly rejected this division of theories, and have essentially made Kant's view *sui generis*.[4]

Given the increasing openness of Kant interpreters to questioning deontological readings of Kant, and their desire to make more room in the theory for a robust view of the virtues, why haven't Kantians made an explicit move toward a virtue theoretic interpretation of Kant? I believe one thing that is holding Kantians back is the suspicion that there are too many theoretical obstacles to developing such an interpretation, that virtue theory is somehow structurally unsuited for capturing Kant's insights.

However, virtue theory has a lot to offer Kantians. Kantian ethics has been attacked not only by those who wish to see emotions and character play a larger role in our picture of the morally good agent, but

[1] See for example Sherman (1997), Herman (1996), Korsgaard (1996).

[2] See for example Williams (1981a, 1981b), Sedgwick (1990).

[3] David Cummiskey (1996) makes the surprising move of arguing that Kant is a consequentialist.

[4] Guyer seems to think Kant's view combines elements of teleological and deontological views. Wood resists any classification of Kant's view. In informal discussion, Christine Korsgaard has also expressed skepticism about the usefulness of traditional classification schemes for interpreting Kant's ethics.

by many who are skeptical about the project of grounding morality in universal reason at all. Can reason itself tell us what is moral and motivate us to do it? What are we to say to the presumably rational person who understands what morality requires but who rejects morality as reason-giving for him? Isn't all reasoning really means-end, figuring out how to satisfy our desires and interests?

Deontology is at a disadvantage in responding to these attacks. If deontological theories make principles of action foundational (I will say more about this below), then there is nothing more to say to someone who rejects moral principles; we cannot ground those principles in something further that they promote, and then it is very difficult to explain their normativity, to explain why they are binding on all of us. Kantians try to demonstrate the rational inescapability of the categorical imperative (CI), but their arguments have not won the skeptics over.

Yet virtue theorists have long been working on this very problem, on showing how moral principles can be binding on us without being reduced to means-end principles, and the structure of virtue theory has been important in making this argument work. Most virtue theorists have also worked hard on avoiding a collapse into consequentialism, something very important to the Kantian. Thus, Kantians should consider the possibility that a virtue theoretic interpretation of Kant would help them address some of the most powerful criticisms of Kantian ethics. Ultimately, I think virtue theorists would also benefit from incorporating Kantianism and relying on a conception of flourishing as a rational agent rather than as a human being, but I will not discuss that in this paper.

In order to clear some space for the project of bringing Kantian ethics and virtue theory together, I am going to examine three of the most apparent obstacles to developing a Kantian virtue theory, and show how they can be overcome. In the process, I will argue both for a particular understanding of what virtue theory requires and for a somewhat unorthodox, but fruitful, interpretation of Kant's ethical writings.

(A) Preliminary definition of virtue theory

In order to determine whether Kantian ethics can be developed into a virtue theory, we must first have a clear understanding of what virtue theory is. One of the best articulations of what is distinct about virtue theory is given by Gary Watson (1990), who argues that virtue theory should be distinguished from other types of theory in the same way that deontological and consequentialist theories are distinguished from each

other.[5] The standard distinction between deontological and consequentialist theories is that a deontological theory makes the right prior to the good, while a consequentialist theory makes the good prior to the right. 'The right' here is defined as a conception of rules or laws for how to act. These could be direct commands (e.g., "Do not kill" and "Help others"), or they could be procedures for how to decide how to act (e.g., "Act in a way that other rational beings could not object to"). 'The good' is defined as a conception of the value of states of affairs. Goodness as it is used here is meant to evaluate the outcomes of actions, as when we say that, other things being equal, it is better to feed ten hungry people than to feed five.

Deontology and consequentialism differ in which concept they make prior to the other. This priority can be seen in two different features of the theories. First, it affects the evaluation procedure that each theory uses. For a deontologist, when we evaluate an action's moral status, the action's rightness or wrongness takes priority over the goodness or badness of the action's outcome. To take a standard example, suppose a runaway trolley is going to hit five people trapped on the track ahead, and the only way to stop the trolley is to push a large man off a bridge so that he falls to his death in front of the trolley. For the deontologist, even if pushing the man will save five lives, maximize happiness, or otherwise create a better state of affairs, doing so might still be forbidden because it is wrong in various ways –unfair, against duty, and so forth. By contrast, for a consequentialist, if the action would truly create a better state of affairs (considering all the ramifications of such an action), then that overrides the action's seeming unfairness, and the action should be done.

Second, and more importantly, these theories differ in which concept they give *explanatory* priority to. For the deontologist, a conception of the right is foundational. This conception can be formulated and defended

[5] By relying on Watson's way of distinguishing theories, I am committing myself to locating the difference between theories in their explanatory structure; theories are identified by the concept to which they give explanatory primacy. Some deontologists may be unsatisfied with this way of defining deontology because it does not seem to leave room for more subtle deontologies that emphasize the importance of moral education, moral judgment, and so forth. Instead, it characterizes deontologies as grounding all of moral life in principles of right. Still, I think it is very useful to trace the order of explanation in a theory. What is the source of normativity in the theory? Ultimately, a careful approach to this question may reveal that some "subtle deontologies" are really virtue theories. Unfortunately, I do not have space to discuss this here; I will simply assume that Watson's characterization of theories is a useful one. I am grateful to a reviewer for pressing me on this.

independently of a conception of the good, so the deontologist does not defend her principles of right action on the grounds that they are the principles that will lead to greatest happiness. The strictest deontologist will explain the *moral* value of happiness and other good outcomes by appeal to the right. Those things are good that would be valued by right-acting people, so the happiness of a dutiful person is valuable. Those states of affairs are good that result from fair and just choices, so an unfair situation is not good, regardless of how much over-all happiness it contains. A less stringent deontologist may have an independent notion of the goodness of happiness, but claim that it is *constrained* by the right; in other words, the good may be pursued within the constraints of the principles of right.

By contrast, the consequentialist straightforwardly explains the rightness of actions in terms of an independent account of the good. If the good is happiness, then the right action is the one that maximizes happiness, or that conforms to a set of rules that would maximize happiness if followed. For the consequentialist, a judgment about what "ought" to be done, where that judgment is not ultimately derived from the goodness of the action's results, is a mysterious sort of judgment that offers no rationale for the action it recommends.

Notice that one important concept has been left out of this contrast between consequentialism and deontology: the concept of the good person. When we discuss morality, we are of course concerned with the rightness or wrongness of an action, as well as with the good or bad consequences of an action, but we are also very concerned with the character of the agent: is she a good person? Is there a pattern to her actions that reveals a settled character trait?

Both deontologists and consequentialists give this concept of the good person, or of moral worth, a secondary role in their theories. For the deontologist, a conception of moral worth, like a conception of the good, can be explained by the grounding conception of the right. The virtuous person is the person who knows the rules of right action and is motivated to follow them. For example, in an early paper John Rawls (1989) gave a deontological reading of Kant, arguing that for Kant the good will "is constituted by a firm and settled highest-order desire that leads us to take an interest in acting from the moral law for its own sake."[6] The goodness of the good person is explained by her attachment to the moral law.

[6] This is only meant as an example of what a deontological reading of Kant would sound like. The full view of Kant that Rawls defends in a number of writings is much more subtle than this one quotation indicates. I thank a reviewer for pointing this out to me.

Similarly, the consequentialist uses her grounding conception of the good to develop a conception of moral worth. Traditional utilitarians, for example, define the good person as the person who values aggregate happiness and acts to maximize happiness. Character utilitarians, on the other hand, give a more elaborate account of the virtues as those character traits that would maximize happiness if universally held, even if the traits do not necessarily lead the virtuous person to act to maximize happiness in particular situations. Though more complicated, this kind of theory still defines and explains virtue in terms of the prior conception of the good.

What would happen if we made the concept of moral worth primary instead? Such a theory would have to articulate a conception of virtue, of what makes a good person, independently of any prior conception of the right or the goodness of states of affairs. This is trickier than it sounds. It would not be enough to make a central place in one's theory for an account of the virtues. If that account of virtues relies on explaining the virtues as desires to act in accordance with independently defined moral laws, the account will end up being deontological. Although Aristotle is usually seen as an early virtue theorist, some have tried to read him in this more deontological way.[7] On the other hand, if the account of virtues defends the virtues as those traits that lead to an independently understood good, such as a happy life or a happy society, the account will be consequentialist. Some have read Aristotle this way as well.[8]

What the virtue theorist needs is an independently justified conception of what it is to be a good person: to function well as a human being, an agent, or a rational being, depending on the particular theory. The virtue theorist then develops a conception of right action and good outcomes grounded on this conception of moral worth. The most straightforward way to do this is to define right action as what the virtuous person would do, and good outcomes as those the virtuous person would pursue (Watson 1990, p. 455).[9] To return to the earlier example, the runaway trolley case would be evaluated by giving an account of what

[7] Christine Korsgaard (1996) seems to be moving toward this kind of view.

[8] See for example Cooper (1975), Santas (1996), Williams (1995).

[9] This is not the only way one might connect the concepts of virtue and the right. Michael Slote (1997) has argued that there are two general types of virtue theory: those that make choiceworthiness by a virtuous person both necessary and sufficient for an action's rightness, so that a nonvirtuously motivated action could still be right, and those theories that make choiceworthiness by a virtuous person only necessary, but not sufficient, because they hold that a right action must also be chosen in a virtuous way.

the virtuous person would do in the situation, or what the virtues in general require in this kind of situation. If a virtue theorist wanted to defend pushing the large man, she could do this by giving an account of why benevolence requires such an action, and an account of why benevolence understood in that way is a genuine virtue. She would also need to say something about why the virtue of justice does not conflict with benevolence in this situation (or at least why it does not override benevolence in this situation).

(B) The first obstacle for a Kantian virtue theory: *Tugend*

Although we only have a preliminary definition of virtue theory at this point, this is enough to generate an immediate problem with developing a Kantian virtue theory. Clearly, a virtue theorist is going to need a robust theory of the virtues, because she explains the rightness of right actions by appealing to what the virtuous person would do. That wouldn't be much of an explanation without a good account of what makes the virtuous person virtuous. So we might think that the first place to look in developing a Kantian virtue theory is at Kant's own theory of the virtues. Yet when we look at that theory, it quickly becomes clear that Kant defines the virtues in terms of a prior principle of right, and thus it looks as though Kant could not possibly be read as a virtue theorist.

This obstacle is nicely brought out in an early article on Kantian virtue ethics by Robert Louden. Louden argues that Kant's moral theory occupies a middle ground between virtue theory and deontology because the theory makes a central place for virtue, yet virtue itself "remains conceptually subordinate to the moral law" (1986, p. 484). Louden bases his argument on Kant's own discussion of virtue (*Tugend*), particularly in the *Tugendlehre* or *Doctrine of Virtue* (DV).

Kant's discussion of virtue in the DV does seem to support this line of thought. Kant defines virtue as "the strength of man's maxims in

---

Slote refers to the first type of virtue theory as act-based and the second type as agent-based. In what follows, when I refer to virtue theories, I will generally have act-based theories in mind. However, my arguments should work for agent-based theories as well. What I require of a virtue theory is that it make choiceworthiness by a virtuous person necessary for an action's being right, and both types of virtue theory meet that requirement. There is a further question of whether Slote's distinction is a helpful one, which I address in (2009). I am indebted to a reviewer for pressing me on this issue.

fulfilling his duty" (6:394).[10] The strength that Kant is speaking of is a kind of strength of will in overcoming natural inclinations, and the greater the obstacle presented by natural inclinations, the greater virtue is displayed in acting on a good maxim.

Kant goes on to distinguish between this general concept of virtue and the particular virtues themselves. Virtue as a general concept is simply "the will's conformity with every duty, based on a firm disposition" (6:395). When considered in this way, virtue is "merely one and the same" (6: 395). So in this sense there is only one virtue: strength of will in acting as one ought. However, Kant adds, we can also think of the particular ends that a good person ought to have, and these correspond to particular virtues. Kant's theory of the virtues, then, seems pretty straightforward: there are various particular virtues, or excellences of character, corresponding to particular ends that reason requires us to set. And then there is a general excellence of character, a sort of master virtue, consisting of strength in following duty and overcoming our inclinations.

What is noteworthy in this theory of the virtues is that both virtue as a general concept and the particular virtues are defined in terms of a prior conception of the moral law. As Louden puts it, "since … virtue is defined in terms of conformity to law and the categorical imperative, it appears now that what is primary in Kantian ethics is not virtue for virtue's sake but obedience to rules" (1986, p. 478). On the basis of this, Louden goes on to argue that Kant's theory falls between virtue ethics and deontology, but Louden is working with a less stringent distinction between the two types of theory. On my use of the terms, which follows Watson, Kant at this point cannot be read as a virtue theorist at all, but only as a deontologist, because a conception of right action (i.e., the moral law) has explanatory primacy.

However, this inference is too quick. If we turn straight to Kant's discussion of 'virtue' to determine if he is a virtue theorist, we assume that he is using the word 'virtue' in the same way we use it when we are trying to distinguish virtue theories from other types of theories. But that assumption is questionable. As we have seen, for the purposes of distinguishing types of theory from each other, we need the concept 'virtue' to mark out considerations about what it is to be a good person, or to have moral worth as a person. This is what enables us to see the important

[10] References to Kant's works will use the volume and page numbers of the German Academy edition of Kant's *Gesammelte Schriften*, ed. the Royal Prussian Academy of Sciences (Berlin, de Gruyter, 1900-), which can be found in the margins of most translations.

differences between virtue theory and deontological or consequentialist views. In other words, what's distinctive about a virtue theory is not that it makes a list of traditional character traits central. Rather, a virtue theory makes a conception of *functioning well* primary. Often this conception will be spelled out in terms of a list of traditional virtues, but it need not be. If the theorist is skeptical about traditional moral psychology (as Kant surely is), he or she could develop a radically different moral psychology. What's critical is that some conception of acting well, or functioning well as an agent, is primary, rather than derivative from an independent rule of right action.

So at this point the question for us is whether Kant himself is using the word 'virtue' to mark out the concept of the good person, and then filling out that concept with his definition of *Tugend*. It is easy to think that he is, especially if we are already operating under the assumption that Kant is a deontologist. The classic deontologist will simply define the good person as the one who has the proper attachment to the principles of right. Virtue will then consist primarily of a set of motivations to act on these prior principles. Kant might seem to be saying exactly this when he says virtue is strength in following the moral law.

But a deontologist who takes this sort of view has to have a moral psychology that is radically different from Kant's. For such a deontologist, knowledge of the right is conceptually separate from the motivation to act rightly, so we can always conceive of a vicious person who grasps the principles of right but lacks the virtuous person's motivations to follow them. Clearly that cannot be Kant's view. It is a central component of Kant's moral psychology that the moral law is intrinsically motivating and that morally worthy action is not motivated by any desires separate from the respect generated by the moral law itself.

Why then does Kant speak of virtue as a strength of will in following the moral law, as though being a good agent requires two separate things –knowledge of what's required and a motivation to do what's required? To answer this we need to look at the role virtue plays in his overall theory. When Kant divides the *Metaphysics of Morals* into the *Doctrine of Right* and the *Doctrine of Virtue*, he follows an inherent distinction between two ways morality can command us: a moral law can tell us to do some action, period (e.g., to keep a promise). Or morality can tell us to perform some action from a particular motivation (e.g., to keep one's promise out of respect for the moral law and not for any ulterior motive). Kant says, "Ethical lawgiving (even if the duties might be external) is that which *cannot* be external; juridical lawgiving is that which can also be external" (6:220). In other words, ethical commands

(given in the *Doctrine of Virtue*), may have to do with external actions (such as keeping a promise), but also contain a command to perform the action from a particular motive, and since internal motivations cannot be coerced by external forces (nobody else can make me be moved by duty, or make me keep my promise because I recognize the reason-giving force of promising) these sorts of commands cannot be the concern of coercive lawgiving by the state. However, the bare action of meeting contractual obligations from whatever motive can be externally coerced, so laws regarding contracts are also discussed in the *Doctrine of Right*.

So the DV is primarily concerned with that part of morality that commands us to be motivated in certain ways and to set certain ends for ourselves. This means that there is already a focus in the DV on motivation as a distinct subject of discussion. The moral law itself is already presumed to be proven and understood, and now the focus is on what ends or motives it commands us to have. This leads to a further narrowing of the topic. Because the concern of the DV is with what motivations we *ought* to have, the DV must be addressed to those who have competing motivations. Morality is commanding us to act on the motive of duty *as opposed to* other incentives, and such a command can only be cogent for beings subject to other incentives, namely, humans. So, while a purely rational being will act in the way the DV requires, and will have the ends the DV discusses, the DV is not directed to such a being. A holy will cannot be commanded to constrain her desires and ulterior motives because she does not have any.

So in the DV, Kant needs a concept that will pick out the particular orientation of the will that the duties of the DV require. The orientation is one that humans in particular need –strength or fortitude– and Kant uses the word 'virtue' to mark out this capacity. Kant is clear, though, that this strength of will is not just another inclination within us, battling with the other inclinations and winning when it is strong enough. If our actions were guided by such an inclination, they would clearly be heteronomous and thus not morally worthy. So what is this strength of will?

We get a clue early in the *Metaphysics of Morals*, in Kant's discussion of moral anthropology. He argues that a moral anthropology, which would detail the particular helps and hindrances humans are subject to in acting morally, must not precede an actual metaphysics of morals. If one worked out the anthropology first, focusing on human incentives to be moral, "one would then run the risk of bringing forth false or at least indulgent moral laws, which would misrepresent as unattainable what has only not been attained just because the law has not been seen and presented in its purity (*in which its strength consists*)…"

(6:217, emphasis added). I take it that Kant means that even when we are considering various ways we might encourage people to be moral and remove temptations to be immoral (for example, in moral education), we should not think of ourselves as adding or enforcing separate incentives to act as duty requires. In the end, what moves a person to be moral (and not just follow the letter of the law) has to be perception of the moral law itself. The more clearly the law is perceived, the stronger one's will is in following it.

Yet we do not want to take the perception metaphor too far. It is not as though the moral law is there to be grasped by any objective observer regardless of that person's motivations. If the moral law only moved us by generating a separate impulse to obey it, we could imagine someone who was deficient in motivation but still able to grasp the law. It would be a contingent matter whether a person was moved to be moral.

What Kant really means to say is clearer in the *Critique of Practical Reason*. In the third chapter, "On the Incentives of Pure Practical Reason," Kant explores the issue of how a law of practical reason can be motivating for the human will. We know that for a human will all action must be motivated in some sensible way, or else we wouldn't be able to explain the action –it would look like a random event. So the moral law, though objective, must have a subjective influence on the will if reason is practical at all. Kant says:

> There is here no *antecedent* feeling in the subject that would be attuned to morality: that is impossible, since all feeling is sensible whereas the incentive of the moral disposition must be free from any sensible condition. Instead, sensible feeling, which underlies all our inclinations, is indeed the condition of that feeling we call respect, but the cause determining it lies in pure practical reason… *And so respect for the law is not the incentive to morality; instead it is morality itself subjectively considered as an incentive*… (5:75-6, second emphasis added).

Here Kant is struggling to identify the connection between the moral law and the will in exactly the right way. He knows that in order for humans to act on a law or reason, they must have some kind of motivation; in other words, the idea of a completely external reason for action makes no sense. Yet, if the moral law can only move us through natural desires, then all commands of morality will be hypothetical imperatives, and only those who have the right desires will be moved to do what's right. Such commands are too contingent, so cannot count as *moral* commands. So Kant has to say that the moral law gives rise to a very particular kind

of motivation: respect. Respect *is* a feeling that moves us (assuming pure reason can be practical for us), but it is not logically prior to the moral law. Rather, it follows from the moral law. Yet it is not simply a separate motivation caused by perception of the law, because then moral commands would be contingent on that causal connection going right in particular cases. Really, respect *is* the law, considered subjectively (that is, considered as a practical and motivating command). So perception of the law cannot actually be separated from motivation to follow it.

This explains both why Kant is so interested in virtue as a kind of strength of will, and why he defines virtue in terms of the moral law. In DV Kant wants to focus on that part of morality that is internal and cannot be commanded externally, namely, on how we should be moved as moral agents. The primary moral feeling is respect, whose effect is felt in the way the agent comes to see her other inclinations as insufficient for justifying action. Respect is felt in the striking down of the pretensions of our inclinations to give us reasons for action directly. So respect is a kind of strength, a strength in distancing oneself from the pull of those inclinations. Yet this respect is not separable from the moral law itself. It does not just function as one desire among others, winning by its strength. It has to be understood as a response to the moral law, or, better, as the very way the moral law shows up for us as sensibly influenced creatures. Kant uses the word virtue to mark out the motivational effects of the moral law on *human* agents. The more clearly we see the law, the stronger is our distancing from our other desires.

This does not seem to be the way the word 'virtue' is employed by contemporary ethical theorists. There we are concerned with virtue as marking out the good agent in general. We usually assume such an agent will be human, but the primary concern is with good agency in general. The closest term to that in Kantian ethics is the good will, which all rational agents can have. He uses virtue to discuss the particular way the good will appears in sensible creatures like humans. And his discussion of the virtues is clearly focused on the particular ends that the moral law commands sensible creatures like humans to have.

This means that if we want to evaluate the prospects of developing a Kantian virtue theory, we will be misled if we look first at Kant's own use of the word 'virtue'. Our question of whether Kantianism is compatible with virtue theory needs to be pushed back to a question about the good will itself. If the good will is Kant's conception of moral worth or good agency, then is that conception itself defined in terms of a prior moral law? And if not, is there any hope of finding or developing a Kantian theory of the virtues of the good-willed agent in general (as opposed to the

"virtues" of humans in particular)? Could we develop a substantive enough theory of the virtues to explain why the good-willed agent is good, and why her actions are right?[11]

(C) The second obstacle for a Kantian virtue theory: codifiability

Refocusing the project this way immediately leads us to another obstacle. The obstacle is that the good will itself seems to be defined in terms of a prior moral law, so we have even more reason to read Kant as a deontologist rather than a virtue theorist. After all, isn't the good will defined as the will that acts on good maxims, out of respect for the moral law? Doesn't a good-willed agent bring her maxim to the categorical imperative, an apparent principle of right action, and then accept or reject her maxims depending on whether that principle can endorse them?

There are really two obstacles here that need to be separated. The first is that the very fact that Kant articulates a specific principle of right action, thus codifying the content of morality, might seem automatically to disqualify him as a virtue theorist. This is because the thesis of noncodifiability has come to be associated with virtue theories, and one of the primary motives for moving towards virtue theories is the loss of faith in the deontologist's project of codifying moral principles (Watson 1990, p. 454). I will show in this section how the Kantian virtue theorist can surmount this obstacle. However, even after the codifiability issue is resolved, there is the further problem that the resulting view still seems to make the categorical imperative logically prior to the good will rather than the other way around. I will address that problem in the next section.

The codifiability issue is complicated because the discussion of noncodifiability by virtue theorists has been somewhat ambiguous, and it is difficult to evaluate how committed a virtue theorist needs to be to noncodifiability. This needs to be clarified before we know whether Kant's codification of a principle of right stands in the way of developing a Kantian virtue theory.

There are two different levels at which an ethical theory can endorse noncodifiability. The first level is in the theory's account of specific moral precepts, rules, and guides for action. A theory can hold either that there are articulable general principles that can guide action in particular cases, or that particular cases are always too particular and will need to

[11] Of course, we will also have to explain how that conception is linked to the more familiar human virtues as well.

be decided by a sensitive person in the situation. Of course, this is oversimplified; most theorists will lie in the middle, believing we can make some generalizations, but that there will be exceptions when we apply our principles to the real world, and even the staunchest rigorist will admit that sensitivity and training are required in order to apply moral principles correctly. However, it is still useful to distinguish two general types of theorists –- those whose theories attempt to codify morality into fairly general principles, such as the Categorical Imperative or the Principle of Utility, and those who resist any such codification and stress the need for moral perceptiveness in actual situations. Aristotle is a good example of the latter type of theorist. I will call the point of contention here *the codifiability of moral principles*.

However, a very different sort of codifiability is sometimes discussed by ethical theorists, most notably by John McDowell. In several important articles, McDowell raises skeptical arguments about the codifiability of morality itself.[12] His arguments attempt to show that even if we are able to articulate general moral principles, these principles will not be intelligible independently of the moral person's outlook and way of life. These arguments attack the widely-held picture of the virtuous person as having two logically independent mental states: knowledge of a rule (and the cases that fall under it) plus a desire to follow that rule. (Or for those who are skeptical about the virtuous person's knowledge being capturable by a set of rules, the common view is that we should at least be able to separate knowledge of what morality requires in a particular case from a desire to be moral and act on that knowledge.) Even if these mental states can't really be psychologically prized apart, the idea is that we should keep them conceptually separate in order to maintain the objectivity of the cognitive state and the motivating power of the appetitive state.

If this picture is accurate, it is logically possible for there to be a person who has the same knowledge as the virtuous person, but who lacks the virtuous person's motivations and responses. This implies that virtue requires both knowledge and a set of desires, and now the door is open for both the Humean and the non-cognitivist to argue that only those with the desire to be moral truly have a reason to be moral.

In opposition to this, McDowell wants to support the Socratic thesis that virtue is knowledge, and he does this by repudiating the idea that the virtuous person's over-all knowledge can even conceptually be

[12] See especially "Virtue and Reason," "Non-Cognitivism and Rule-Following," and "Are Moral Requirements Hypothetical Imperatives?," in McDowell (1998).

separated from her motivations. He argues that an outsider to the virtuous person's way of life could not match the virtuous person's knowledge, since her knowledge consists of her normative stance itself.[13] This makes the virtuous person's moral knowledge not fully graspable by those outside the practice of virtue, and it makes it possible to support the claim that knowledge of the good is intrinsically motivating and does not require us to posit a separate set of desires that would give an agent reason to be moral. This view is distinct from the view that moral knowledge cannot be codified into principles, since one could hold that the virtuous person's knowledge can be articulated by a moral code, but that fully grasping that code requires the acquisition of virtue; on the other hand, one could hold that the virtuous person's knowledge cannot be codified into principles, but that her responses to individual situations can be separated into a neutrally describable piece of knowledge and an emotional response to that knowledge. So I'm going to refer to the issue at stake here as *the independent intelligibility of moral knowledge*.

Now let us explore the extent to which a virtue theorist must be committed to either the noncodifiability of moral principles or the lack of independent intelligibility of moral knowledge. It is fair to say that Aristotle is at least skeptical about the codifiability of moral principles. But could a virtue theorist go in for codifiability? It seems that she could, as long as she maintained the proper relation between her conceptions of the right and of virtue. She could argue that the right is what would be done by the virtuous person, and as a matter of fact the virtuous person will follow the Principle of Utility, or the Categorical Imperative, or some other principle. The trick here is in how she makes this argument. If she argues independently that the principle is right, and that this explains why the person who follows it is virtuous, the view will really be deontological or consequentialist (depending on how the principle itself is grounded). To spell out a truly virtue theoretic view, she needs to provide an independent argument that a certain kind of person is virtuous, and then show that a moral principle or principles in fact do describe the virtuous person's choices, but have no independent status. They are the correct principles *because* they describe the virtuous person's choices.

Again, though, it is trickier to do this than it sounds. Even if a theorist claims to be making a conception of "functioning well as an x"

[13] The trick for McDowell is to show how such a stance can count as cognitive and not a mere projection of desires, but that argument is too involved to discuss here.

foundational, if there is nothing more to that conception of functioning well than just that it accords with a certain principle, then the resulting theory doesn't deserve to be called a virtue theory. A true virtue theory needs a robust account of what it is to be functioning well as an *x*, an account that is not fully described by a principle of action. So a virtue theorist can endorse the codifiability of moral principles, but only as long as she does not see such a codification as fully describing virtuous action.

Regarding the independent intelligibility of moral knowledge, I believe a virtue theorist can go either way on this issue as well. Obviously, a McDowellian virtue theorist would reject independent intelligibility. On the other side, Philippa Foot is a good example of a virtue theorist who embraces independent intelligibility.

In recent work, Foot defends a moral theory that is clearly virtue theoretic in structure, and that sees moral facts as intelligible to any scientific observer (Foot 1995, 2001). To give a quick sketch, on Foot's view the virtues are Aristotelian necessities for our species: characteristics that are necessary for the flourishing of our species. For example, humans need benevolent and just members of society in order for human society to flourish, so benevolence and justice are Aristotelian necessities for us. Interestingly, Foot makes an explicit analogy between judgments about what is good for the human species and judgments about what is good for other animal species. Just as it is a plain matter of fact that there is something wrong with an owl that cannot see in the dark, or with a lioness that ignores her cubs, it is a plain matter of fact that there is something wrong with a person who is not moved by considerations of justice. In each case, we can judge from a loosely scientific point of view that this is a defective member of the species. Because Foot sees judgments about what makes a virtuous human as quasi-scientific judgments, I think it is fair to say she takes these judgments to be independently intelligible – intelligible to any competent observer of our species, whether that observer is virtuous or not. The virtuous person, then, would be someone who both recognized what the species needed to flourish and was motivated to be a flourishing member of the species.

To summarize, we have four categories into which virtue theorists can fall: (1) Theorists who take moral principles to be codifiable, and who take this codification to be independently intelligible. (2) Theorists who do not think moral principles are codifiable, but who do take specific moral knowledge to be independently intelligible. (3) Theorists who think moral principles can be codified, but that this codification will not be independently intelligible. (4) Theorists who think moral principles

cannot be codified, and that even specific moral knowledge will fail to be independently intelligible. Foot is a good example of type (2), and McDowell is a good example of type (4).[14]

This classification scheme opens up new options for developing a Kantian ethics. I believe that the fact that Kant is a strong codifier of moral principles has biased interpreters toward a deontological reading of him, rarely seeing a virtue theoretic interpretation as an option. But as we have just seen, a virtue theorist can still maintain codifiability. As we have also seen, endorsing codifiability does not commit a theorist to holding that the resulting codification is independently intelligible. So if we attempt to develop a Kantian virtue theory, it will be important to determine whether a Kantian needs to hold the thesis of independent intelligibility or not. Thus, if we assume that the Kantian will endorse codifiability, we still have two issues to decide: first, can the resulting theory have the structure of a virtue theory, even though it provides a codification of moral principles? Second, does the theory need to see that codification as independently intelligible or not? What I am going to argue is that we can develop a Kantian virtue theory of the third type – one that endorses codifiability but rejects independent intelligibility. I think it is most promising to develop a Kantian virtue theory of this third type because Kant clearly leans toward a codification of moral principles, but (as I mentioned earlier) he also has a moral psychology that requires that these principles not be independently intelligible because they need to be intrinsically motivating.

(D) The third obstacle for a Kantian virtue theory: independent intelligibility

This brings us to the third and most difficult problem for developing a Kantian virtue theory. The problem is that the categorical imperative still seems to play a grounding role in Kant's theory. To make any room for a Kantian virtue theory, we need to answer this foundational question: does the CI explain the goodness of the good will, or does the goodness

[14] It might seem to be a problem that there are no clear examples of categories (1) and (3). I think this can be explained by the fact that (1) has all the difficulties of deontology in grounding the reason-giving force of its codified principles, with few of the benefits of virtue theory. I will argue that Kant can be read as a theorist of type (3) but that this has not been recognized because of his deontological-sounding language.

of the good will explain the rightness of the CI itself? An early attempt to answer this question in favor of a virtue theoretic reading is given by Onora O'Neill in "Kant After Virtue" (1984). There, O'Neill argues for the view that the Kantian agent's maxim should be understood as the more general "*underlying intention* by which the agent orchestrates his numerous more specific intentions" (1984, p. 394). Her reason for moving to this view of maxims is that the CI requires that we act on maxims that could be universally acted on, and underlying intentions are far better suited to this universalization test than specific intentions are. For example, the specific intention to lie to Jane Doe on a particular date would seem to pass the CI test, but the underlying intention to deceive in order to get out of trouble would not.

O'Neill thinks this view of maxims naturally leads to a virtue theoretic reading of Kant. Since maxims are underlying intentions, "to have maxims of a morally appropriate sort would then be a matter of leading a certain sort of life, or being a certain sort of person. The core of morality would lie in having appropriate underlying intentions rather than in conforming one's actions to specific standards" (1984, p. 395). Furthermore, this conception of how to live as a certain kind of person would itself ground judgments of right and wrong: "It is clear enough that for Kant the categories of virtue …are more fundamental than the categories of right… For his definition of right action is that it conforms in (at least) outward respects to action that is done out of a morally worthy maxim" (1984, p. 396).

So the picture seems to be this: Kant defines right action as what would be done by someone of good will, someone acting on a good maxim. In addition, to act on a good maxim is not to conform one's behavior to independent rules of right, but to be a certain kind of person with certain kinds of underlying intentions. Thus, right and wrong are explained by a conception of the morally worthy agent, and we seem to have a virtue theory.

But this is too easy. It may be that the rightness of particular actions is explained by what a good-willed agent would do, but for all O'Neill has told us that conception of the good-willed agent itself is still explained in terms of the categorical imperative. As Kant says, the good-willed agent is the one whose maxim is good and cannot be bad, and a maxim is good if it conforms to the categorical imperative. The CI looks like a paradigm example of a principle of right: it sorts maxims into permissible and impermissible. Though it does not sort actions themselves, it does sort principles of action, and seems ideally suited to be a code an agent could consult when she is trying to decide how to act.

If Kant's view is that the principle explains why a good-willed agent is good (because her maxims conform to this principle), that view certainly deserves to be called deontological. If we want to read Kant as a virtue theorist, we will have to do something more radical than O'Neill has done in reading maxims as underlying intentions. We will have to reverse the order of explanation between the CI and the conception of the good-willed agent, so that the normativity of the CI itself is *explained by* the prior conception of the good will. And given the textual evidence cited so far, this looks like a hopeless task.

However, a closer examination of Kant's texts reveals more ambiguity than one might expect. In fact, one can find clear evidence that Kant was drawn to two claims: first, that the categorical imperative is not meant to be a self-sufficient principle, but that it is simply a codification of a complicated and substantive ideal of rational agency, and that its rightness is explained by that ideal. Second, that the normativity of the categorical imperative cannot be grasped from outside the perspective of a person already committed to the value of rational agency, and so the CI is not independently intelligible. Clearly, Kant was drawn to the project of articulating a decisive formula for making moral judgments, and he was committed to grounding morality in universal reason and not in human nature, but we should not be misled by the resulting deontological language into assuming Kant's view is actually best developed as a deontology.

(i) The status of the categorical imperative

Evidence for the idea that Kant did not mean the categorical imperative to stand alone as a grounding principle of right can be found primarily in the *Groundwork* and the *Critique of Practical Reason.* Evidence from the *Groundwork* largely comes from its over-all structure. Section One is explicitly meant to start from our common understanding of morality and move toward its underlying principles, and the main concept Kant starts from is our ordinary understanding of the good will. Throughout the section, the essential elements of morality, which will later be encapsulated in the categorical imperative, are drawn out from our ordinary understanding of what a morally good agent is like and how she goes about making decisions. In particular, Kant shows us that our ideal moral agent acts from a special motivation: she does the right thing because it is right, not solely in order to satisfy her own desires. In other words, she does not take her own desires as reason-giving on their own. Using a series

of counter-examples, Kant shows us that we do not actually think that any desire or any particular end can be sufficient for morally worthy action. Even the most benevolent desires do not give a person good will if that person is only accidentally led to acting well. And even the best ends (such as others' happiness) can accidentally be brought about by an evil will. What Kant shows us in Section One is that we are already committed to the view that it is a necessary condition of acting well that a person know what she is doing and that her action flow directly from her own free choice. Since no particular desire or end will guarantee this kind of action, no particular desire or end can be sufficient for morally worthy action. This much we can get from our ordinary intuitions about morality.

Only at this point in the text (toward the end of Section One) does the philosophical work begin. Once we know that the good will cannot consist merely of a certain set of desires or ends, we require philosophical analysis to explain what the good will must consist of. What are the necessary and sufficient conditions for morally worthy action? What ties all good-willed actions together? Here Kant claims that we will need to find a principle or law that the good will follows, and thus begins the deontological-sounding language. But Kant has a very good reason to stress the need for a principle here: either the good will is acting in a principled way, or it is acting randomly. The good will must act for some *reason*, on some consideration that objectively ties all its actions together as good, or it acts for no reason. But action performed for no reason at all would only be accidental, and we have already seen that such action would not be morally good in our ordinary understanding of morality. It is simply a necessary part of our understanding actions as rationally guided that we have to see them as performed for some reason, so there has to be some principle or general consideration (or at least some set of them) that underlies all good actions. However, this does not mean that the goodness of those actions is explained by their conformity to this principle.

So, given that there must be some principle to which we can see the good will's actions as conforming, our philosophical task is to articulate that principle. However, we have got very little to work with because we have seen that no particular desire or end will be sufficient to give us a principle of good willing. For example: *Follow all your friendly, kind, helpful feelings and resist your hurtful, angry, selfish feelings* is not the principle of the good will. Neither is: *produce the greatest amount of over-all happiness*. So Kant's answer is this:

> Since I have deprived the will of every impulse that could arise for it from obeying some law, nothing is left but the conformity of actions as

> such with universal law, which alone is to serve the will as its principle, that is, *I ought never to act except in such a way that I could also will that my maxim should become a universal law* (4:402).

In other words, the only principle that we can see as uniting all the actions of a good will is the principle to act on principle itself. Only action that follows this principle is guaranteed to be performed because it is right and to be freely chosen.

Section Two of the *Groundwork* goes on to give a more philosophical analysis of this intuitively legitimate moral principle, by showing how the principle itself can be derived from the very concept of rational willing. Rational wills can act on two kinds of principles: hypothetical and categorical. As we have already seen from Section I, truly morally worthy action would have to be performed on the basis of categorical imperatives because moral action must be done for its own sake, not for contingent reasons. After distinguishing the two kinds of imperatives, Kant reiterates the argument given in Section One to show that the only possible categorical imperative would be the command to act on maxims that one could will as universal law.

One might expect Kant to stop with this formulation of the categorical imperative (usually referred to as the universal law formulation). After all, he has just shown (twice) that this is the only possible moral principle. But Kant goes on in Section Two to give several more formulations of the categorical imperative. He claims these formulas are simply "three ways of representing the principle of morality" (4:436), so they are not essentially different. Why argue for them, then?

A clue can be found in the transition from the universal law formulation to the next formulation: the formula of humanity as an end in itself. After giving the universal law formulation, and showing how it works, Kant surprisingly says:

> But we have not yet advanced so far as to prove a priori that there really is such an imperative… The question is therefore this: is it a necessary law *for all rational beings* always to appraise their actions in accordance with such maxims as they themselves could will to serve as universal laws? If there is such a law, then it must already be connected (completely a priori) with the concept of the will of a rational being as such. But in order to discover this connection we must, however reluctantly, step forth, namely into metaphysics… (4:425-26).

This is surprising because one would have thought that Kant had already given the argument for the formula of universal law being connected to the concept of rational willing. However, he is not satisfied with the argument he has given so far, and so he goes on to the more 'metaphysical' argument for the formula of humanity:

> If, then, there is to be a supreme practical principle and, with respect to the human will, a categorical imperative, it must be one such that, from the representation of what is necessarily an end for everyone because it is *an end in itself*, it constitutes an *objective* principle of the will and thus can serve as a universal practical law. The ground of this principle is: *rational nature exists as an end in itself* (4:428).

Only now does Kant give the formula of humanity: that we ought always to treat humanity as an end and never solely as a means.

I interpret Kant to be saying something like this: we can get the *content* of the categorical imperative simply by looking at the concept of a categorical imperative, but we cannot get its *normativity* for us as rational beings without examining the concept of rationality as an end in itself.[15] Rationality as an end, as something to be valued, is the ground of the CI's normativity; it is what makes the CI reason-giving.[16] Once we understand what it is to value that capacity as an end in itself, we are led to the "very fruitful concept" of a kingdom of ends, and we get the final formulation of the CI. Thus, the CI itself does not seem to capture fully what it is to function *well* as a rational agent. While a good-willed person's actions will in fact be in accord with the CI and will be performed out of respect for the CI, the CI by itself does not fully explain why this counts as functioning well as a rational agent and why we have reason to aspire to function well as rational agents. Only with the development of all the

[15] By "content" here, I mean only that we can figure out what the CI *says*, what it tells a rational agent to do. At 4:436, Kant says that the first formulation of the CI gives us the *form* of maxims; I don't mean to use the word "content" as a contrast to Kant's use of "form". Rather, telling us what form maxims take just *is* the content I am referring to. I thank a reviewer for pressing me on this.

[16] I take my argument here to be compatible with the argument Barbara Herman makes in "Leaving Deontology Behind." She argues that "[the] successive formulations [of the CI] interpret the arguments of the CI procedure in terms that reveal the aspects of rational agency that generate contradictions under universalization. These interpretations provide the requisite connection between formal principles and value; they show *how* content is derived from the constraint of universal form for willing" (1993, pp. 227-228). I owe a great deal to her reading of Kant.

formulations of the CI, including the fuller account of what they *mean*, do we get a full picture of the value of rational agency.

Further support for this reading of Kant can be found in the *Critique of Practical Reason*. Early on in the book, Kant again makes the argument that the categorical imperative is the only possible law of a free will. In fact, he shows that "freedom and unconditional practical law reciprocally imply each other," and then he makes a very suggestive comment:

> Now I do not ask here whether they are in fact different or whether it is not much rather the case that an unconditional law is merely the self-consciousness of a pure practical reason, this being identical with the positive concept of freedom… (5:29).

Clearly, Kant is in favor of the latter claim, the claim that the categorical imperative simply *is* our consciousness of our own status as free beings, though he does not argue for it here. Here he is mainly concerned to discover how we humans in fact come to recognize our own status as rational beings. Do we start with an experience of freedom and then derive the moral law from it, or the other way around? Since we can have neither an experience of freedom nor an intuition of it, Kant concludes that we must start with a direct cognition of the moral law. He calls this the "fact of reason":

> Consciousness of this fundamental law may be called a fact of reason because one cannot reason it out from antecedent data of reason, for example, from consciousness of freedom (since this is not antecedently given to us) and because it instead forces itself upon us of itself as a synthetic a priori proposition that is not based on any intuition… [It] is not an empirical fact but the sole fact of pure reason which, by it, announces itself as originally lawgiving… (5:31).

So, in a way, a principle of action *is* fundamental: the moral law is simply given to us, and is the source of our consciousness of our own autonomy. But this principle is not an independent rule that we receive from outside ourselves. Rather, our consciousness of it *as a law*, as normative for us, simply *is* our consciousness of our own autonomy. That's why Kant immediately goes on to fill out the concept of autonomy and explain its difference from heteronomy: he wants to show us what is normative about the moral law for us, not only what its content is or what duties it sets out for us. So, he says:

> *Autonomy* of the will is the sole principle of all moral laws and of duties in keeping with them… Thus the moral law expresses nothing other than the *autonomy* of pure practical reason, that is, freedom, and this is itself the formal condition of all maxims, under which alone they can accord with the supreme practical law (5:33).

Because Kant has already shown in the first *Critique* that we cannot have an intuition of our own autonomy, or any cognitive knowledge of it, when he says that our freedom is expressed in the moral law, he cannot mean that the moral law gives us theoretical knowledge of our own freedom. The moral law is not a proposition stating matters of fact. Rather, it is a command with binding force, and it is our apprehension of it *as normative* that is an expression of our freedom. Any being with theoretical reason could understand the words of the categorical imperative, and could even see that some actions are forbidden by it (maxims that are based on conventions, such as making promises, are especially easy to run through the CI procedure), but unless that being had a sense of herself as autonomous, she would not actually grasp the CI as a normative rule.[17] To grasp it as normative is to express one's autonomy. Thus, the CI cannot stand alone as a rule for action; its force is grounded in a certain standpoint from which we express (and value) our own autonomy.

Furthermore, valuing our own autonomy involves much more than having some theoretical understanding of ourselves as free-willed. After explaining the distinction between autonomy and heteronomy in the second *Critique*, Kant sums up his argument as follows:

> This Analytic shows that pure reason can be practical –that is, can of itself, independently of anything empirical, determine the will– and it does so by a fact in which pure reason in us proves itself actually practical, namely autonomy in the principle of morality by which reason determines the will to deeds. At the same time it shows that

[17] Further support for this point can be found in the *Religion:* "[From] the fact that a being has reason [it] does not at all follow that, simply by virtue of representing its maxims as suited to universal legislation, this reason contains a faculty of determining the power of choice unconditionally, and hence to be "practical" on its own; at least, not so far as we can see… Were this law not given to us from within, no amount of subtle reasoning on our part would produce it or win our power of choice over to it" (6:26n). Thus, the moral law has to be seen as an expression of what we take ourselves to be, an ideal we set for ourselves; it cannot be a rule we grasp with theoretical reason and then realize we should obey. I am indebted to an anonymous reviewer for directing me to this very helpful passage.

> this fact is inseparably connected with, and indeed identical with, consciousness of freedom of the will, whereby the will of a rational being that, as belonging to the sensible world cognizes itself as, like other efficient causes, necessarily subject to laws of causality, yet in the practical is also conscious of itself on another side, namely as a being in itself, conscious of its existence as determinable in an intelligible order of things… (5:42).

Clearly much more is involved in being a moral agent than seeing that a particular rule applies to oneself. To be moral requires seeing oneself as in some way independent from causal determination, as able to act in an intelligible and justified way, and not simply in a mechanically explainable way. Kant is not giving us a list of independently justified rules or commandments; he is articulating the entire outlook that underlies our moral behavior. Thus, his conception of ideal rational agency does not seem to be captured by the CI itself. On the contrary, the rationale for the CI seems dependent on the underlying conception of functioning well as a rational agent. This means that a conception of how to live well is doing real explanatory work in the theory. Conceiving of oneself as autonomous is not a matter just of seeing that certain actions are required or forbidden. Rather, it is a stance we can take on ourselves wherein we see our desires as up for question, our selves as separate from them and able to evaluate them, and our agency as an independent force in the world. Taking such a stance involves both cognitive and emotional states, such as believing one has a choice about how to act and feeling guilt about one's own past choices or resentment of others'. Thus, this stance deserves to be called an outlook in the virtue theoretic sense and it is used to explain the normativity of the CI, not the other way around.

(ii) The independent intelligibility of moral knowledge

We still have not settled, though, what kind of virtue theory would best fit Kant's project. Assuming the CI is grounded in a substantive view of the value of rational agency, is that view itself something that can be fully grasped by someone who is not at all moved by it? For example, could an amoralist form a maxim, see that it would be rejected by the CI, even see that acting on the maxim would violate the value of rationality as an end in itself, and still say: "So what?" In other words, is the value of rational agency as an end in itself independently intelligible?

At first glance, the answer appears to be "yes." We might agree with the virtue theorist that Kant grounds morality on a rich view of what it means to act well as a rational agent, rather than grounding it on a simple rule for right action. But surely Kant is trying to offer us a theoretical account of what a rational agent is and why good rational agency follows a certain pattern? Surely we can imagine a sociopath fully understanding the cognitive content of the good-willed agent's value system, but simply having the opposite attitudes toward that content? After all, Kant gives us all the cognitive content of the good-willed agent's moral outlook in a series of nice definitions: the will is practical reason, and practical reason is the capacity to act in accordance with principles. This capacity should be valued as an end in itself; that is, it should never be used solely as a means to any other purpose. Acting in this way expresses our autonomy, which is simply the ability to give laws to ourselves, or freedom. The good-willed agent is the one who values her own and others' autonomy in the right way. Presumably, then, the thoroughly bad-willed agent would be just as capable of categorizing other beings into rational and arational, and would know when an action involved using a rational being solely as a means, but would simply freely choose to treat rational beings as means to other purposes she has because she values those purposes more than she values the rational capacity. The difference between the virtuous and non-virtuous agent seems simply to be a difference in motivation, not in knowledge.

Again, though, the textual evidence is ambiguous. Kant's repeated insistence that he is grounding the moral law in a priori reason, and not in experience or in human nature, certainly makes it sound as though any being capable of theoretical reason can fully grasp the moral law, regardless of whether that being is at all motivated to be moral. However, as I discussed earlier, there is some evidence against this reading in the *Critique of Practical Reason.* In particular, Kant is clear that the moral law is motivating in itself; grasping it already involves being moved by it. Furthermore, there is an important discussion of moral motivation in the *Metaphysics of Morals* that is quite relevant here.

In the *Doctrine of Virtue*, after arguing for the main ends rational agents should have, Kant goes on to discuss the actual feelings human agents will have insofar as they are rational:

> There are certain moral endowments such that anyone lacking them could have no duty to acquire them. They are *moral feeling, conscience, love* of one's neighbor, and *respect* for oneself (*self-esteem*). There is no obligation to have these because they lie at the basis of morality, as

> *subjective* conditions of receptiveness to the concept of duty, not as objective conditions of morality… To have these predispositions cannot be considered a duty; rather, every man has them, and it is by virtue of them that he can be put under obligation. Consciousness of them is not of empirical origin; it can, instead, only follow from consciousness of a moral law, as the effect this has on the mind (6:399).

Kant clearly means for these feelings to be subjective motivations, the kind of thing that can cause action in human beings. But, of course, he cannot ground the normativity of the moral law on them. It is not as though the reason we should act on the moral law is to satisfy these feelings.

Can Kant have it both ways? Can he say that there are certain subjective feelings we must have in order to be put under moral obligation, yet claim that the obligation applies to us objectively and not contingently on our having such subjective feelings? To show how Kant tries to work this out, I want to focus on two of these "moral endowments": moral feeling, and love of one's neighbor.[18]

Kant defines moral feeling as "the susceptibility to feel pleasure or displeasure merely from being aware that our actions are consistent with or contrary to the law of duty" (6:399). So, technically, moral feeling is not a feeling at all, but a susceptibility or disposition to feel other feelings. In particular, it is a disposition to feel a pro-attitude toward moral actions, and a con-attitude toward immoral ones. But if this disposition is one of the subjective conditions "by virtue of [which a person] can be put under obligation," then haven't we simply reduced morality to something with only subjective binding force again? Morality does not even apply to those who are not disposed to feel pleasure at their moral maxims!

Kant gets out of this problem in an interesting way, by making a distinction between kinds of feelings:

> Every determination of choice proceeds *from* the representation of a possible action *to* the deed through the feeling of pleasure or displeasure, taking an interest in the action or its effect. The state of *feeling*… here (the way in which inner sense is affected) is either *sensibly dependent* or *moral*. The former is that feeling which precedes the representation of the law; the latter, that which can only follow upon it (6:399).

[18] I believe my arguments could be applied to the other two moral endowments as well, but space constraints prevent me from discussing them here.

Kant admits that human action must be caused by some kind of feeling or other; as sensible creatures, our actions need to be *motivated*, either through attraction or aversion. But there are two ways we can understand this motivation –either as something that precedes the law, or as something that follows it.

I do not think Kant means for this distinction between motivations to be a psychological one. He is not saying that there is a temporal sequence in our minds from a desire to a law, or from a law to a desire. Rather, he is making a conceptual distinction. When we cite a motivation in order to explain an action (as opposed to a mere behavior or event), we do this both to give a causal explanation and also to give a rational explanation. Mere behaviors can be fully explained by citing impulses, but actions can only be fully explained by showing what the agent took herself to be doing, what point or good she saw in her action. So motivation, understood as something that can explain action, must have some intentional content. It must cite something the agent was trying to do. What Kant seems to be saying here is that there are two general categories into which motivations can fall: either we take the action to make sense to the agent as fulfilling a logically prior desire, or we take the action to make sense to the agent as being morally right regardless of her other desires. When we understand the motivation in the second way, we are genuinely ascribing a feeling and desire to the agent, something that caused her action, but we are ascribing a very special kind of feeling. It is a feeling that can only be understood as motivating once we understand the moral law itself. We cannot see the point of the action from the agent's point of view without understanding the moral law itself.

This account of moral feeling does allow Kant to have it both ways. On the one hand, he can fully admit that all action, moral action included, must be motivated by sensible feeling. If a being lacked the capacity for moral motivation, the moral law would not be obligatory for that being because that being could not truly act on the moral law. But on the other hand, the moral law's binding force is objective because the feeling that motivates moral action is not intelligible independently of an understanding of the point of the moral law to begin with. The moral feeling cannot be a feeling that some rational beings just happen to have and others happen to lack. The moral feeling just *is* the motivation that sensible rational beings have when they try to justify their actions. If Kant's previous arguments for the moral law's being the only possible law of practical reason hold true, then all rational beings must have moral feeling. To lack such feeling would get one off the hook of moral obligation,

but it would also mean one never truly *acted* at all. Of course, Kant has not proved, and cannot prove, that anyone ever actually has grasped the moral law and been moved by moral feeling, but his view clearly seems to be that grasping the moral law would logically entail being motivated by it. So we *cannot* actually conceive of a rational being who fully understands the categorical imperative, but does not care at all about it.

Of course, this might seem to be a trivial point. All we have said so far is that the moral law only applies to those with a *capacity* to be moved by morality. But *of course* that is true. The moral law only applies to beings who are conscious and capable of acting, as well, but there is nothing interesting about that. However, we have only looked so far at one of the moral endowments that Kant thinks are necessary grounds for moral obligation. There are three others, and the most surprising one is love of one's neighbor.

Love of one's neighbor is a subjective feeling, so is not something we can be commanded to have. Thus, Kant wants to distinguish it clearly from benevolence and beneficence, or dispositions to do good things for others, which are traits we can be morally required to cultivate. The particular moral endowment that Kant calls "love of one's neighbor" or "love of man" is simply "*delight* [in another's perfection] … which is a pleasure joined immediately to the representation of an object's existence" (6:402). In other words, it is an immediate pro-attitude connected to our perception of others as autonomous beings, and Kant's view is that if we did not already have this feeling, we would not be able to act morally at all.

Though Kant does not spend much time explaining love of man, the implications of what he does say are quite important. Since he is very clear that we cannot have cognitive knowledge of anyone's status as autonomous, including ourselves, and that we cannot even know that anyone ever actually is moral, what can he mean by saying we must have a feeling of delight in others' perfection? He cannot mean that we must make a cognitive judgment that others are good rational agents before we can act morally, nor that we need to estimate how perfect others are based on the evidence of their good actions. He must mean something closer to the idea that we must have a certain attitude towards others: we must relate to others as fellow rational beings, and love the fact that we can relate to them that way. We must delight in the special kind of relationship that only rational beings can have with each other.

There is nothing terribly controversial about the idea that adult humans can relate to each other in a profoundly different way from how we can relate to infants, animals, and non-conscious things. Only fellow rational beings are held fully responsible for their actions, and only toward

them is it appropriate to feel emotions like resentment and moral admiration.[19] What is controversial is the claim that only by having that kind of relationship with others can we see the point of morality. We already have to see others as having a special status before we can grasp our obligation to act on universalizable maxims.

This reading of "love of man" fits well with the idea that the different formulations of the CI in the *Groundwork* are much more than simple restatements of the same law. The formula of humanity requires a certain way of looking at other people, a way of relating to them as fellow rational beings, and this perspective is necessary in order for us to see the normative force of the CI. The formula of the kingdom of ends, then, could be seen as articulating a social ideal where each person relates to each other person as an equal, rational agent. Only by sharing in that ideal can we understand what universalizability requires, and why it is so important.

One implication of this reading is that it now becomes completely consistent with Kant's view to claim that becoming a moral agent requires a great deal of substantive training in a community of moral agents. Simply ordering someone to follow the CI would not make that person a moral agent or make their actions good. Acquiring the right perspective on her own desires and on the status of other people would be necessary for acquiring moral agency, and this would presumably require much experience. Yet, once we have acquired that perspective we can see that the moral law is objectively grounded and inescapable, so we have maintained its a priori status. The point of putting the moral law into various formulations and arguing for its rational status is to articulate to ourselves what it is we think we are doing when we strive to be moral, and hopefully to improve our own understanding of what morality requires of us. What has explanatory primacy, though, is the underlying outlook itself –the set of beliefs, emotional responses, and ends that drive us to formulate principles of moral behavior.

Thus, it is possible to develop a truly Kantian ethical theory that is virtue theoretic in structure. The trick is not to read the categorical imperative as a grounding principle of right, but as itself grounded in a conception of good agency. And it is vital to see that conception of agency as itself only intelligible to those who are already in the business of trying to function well as agents, because this is the only way to preserve Kant's claim that the categorical imperative is intrinsically motivating. The

[19] Of course, as rational beings-in-progress, children are the recipients of some of these attitudes.

advantage to reading Kant in this way is that it allows us to avoid the problems of deontology and gives us new insight into some of Kant's texts. At the same time, we are able to retain the Kantian insistence that morality is grounded in a priori reason, and that the moral law gives a reason for action to all rational agents that is not grounded in their contingent desires.[20]

(E) Conclusion

I have tried to clear away some of the most obvious obstacles to developing a Kantian virtue theory. The fact that Kant develops a deontological account of the human virtues need not prevent his over-all theory from having the structure of a virtue theory. Furthermore, his codification of the categorical imperative and his insistence that the good will follows the categorical imperative need not force us to read him as a deontologist, especially once we have a clearer understanding of what virtue theory itself requires. Obviously, much more work needs to be done to develop a Kantian virtue theory, especially in developing a fuller theory of the virtues of rational agency, but I hope to have made space for such a project.[21]

References

Cooper, J. (1975), *Reason and Human Good in Aristotle*, Cambridge, MA, Harvard University Press.

Cummiskey, D. (1996), *Kantian Consequentialism*, New York, Oxford University Press.

Foot, P. (1995), "Does Moral Subjectivism Rest on a Mistake?", *Oxford Journal of Legal Studies,* 15 (1), pp. 1-14,

——— (2001), *Natural Goodness,* Oxford, Oxford University Press.

Guyer, P. (2000), *Kant on Freedom, Law, and Happiness*, Cambridge, Cambridge University Press.

[20] A more problematic implication of this reading of Kant is that all rational beings by definition are motivated to be moral and do see the normativity of the categorical imperative, so we cannot actually conceive of a truly amoral rational agent, and radical evil becomes unintelligible. Troubling as this is, it does seem to fit well with what Kant himself says about the incomprehensibility of evil in *Religion*, 6:43-44.

[21] I am very grateful for helpful comments from Lara Denis, John Exdell, Jim Hamilton, Jerry Santas, Gary Watson, Donald Wilson, Sofia Sabatés, Marcelo Sabatés, and several anonymous reviewers.

Herman, B. (1993), "Leaving Deontology Behind", in *The Practice of Moral Judgment,* Cambridge, MA, Harvard University Press, pp. 208-240.
——— (1996), "Making Room for Character", in Engstrom, S. and Whiting, J. (eds.), *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty,* Cambridge, Cambridge University Press, pp. 36-60.
Kant, I. (1797/1991), *The Metaphysics of Morals*, Gregor, M. (trans. and ed.), Cambridge, Cambridge University Press.
——— (1793/1996), *Religion Within the Boundaries of Mere Reason,* di Giovanni, G. (trans.), Cambridge, Cambridge University Press.
——— (1788/1997a), *Critique of Practical Reason*, Gregor, M. (trans. and ed.), Cambridge, Cambridge University Press.
——— (1785/1997b), *Groundwork of the Metaphysics of Morals,* Gregor, M. (trans. and ed.), Cambridge, Cambridge University Press.
Korsgaard, C. (1996), "From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action", in Engstrom, S. and Whiting, J. (eds.) (1996), *Aristotle, Kant, and the Stoics, rethinking happiness and duty,* Cambridge, Cambridge University Press*,* pp. 203-236.
Lara, A. (2009), "Agent-Based versus Agent-Focused Virtue Theories: A Counterproductive Distinction", *Southwest Philosophy Review,* 25 (1), pp. 199-206.
Louden, R. (1986), "Kant's Virtue Ethics", *Philosophy,* 61, pp. 473-489.
McDowell, J. (1998), *Mind, Value, and Reality*, Cambridge, MA, Harvard University Press.
O'Neill, O. (1984), "Kant After Virtue", *Inquiry*, 26, pp. 387-405.
Rawls, J. (1989), "Themes in Kant's Moral Philosophy", in *Kant's Transcendental Deductions,* Forster, E. (ed.), Stanford, Stanford University Press, pp. 81-113.
Santas, G. (1996), "The Structure of Aristotle's Ethical Theory: is it Teleological or a Virtue Ethics?", *Topoi,* 15, pp. 59-80.
Sedgwick, S. (1990), "Can Kant's Ethics Survive the Feminist Critique?", *Pacific Philosophical Quarterly*, 71, pp. 60-79.
Sherman, N. (1997), *Making a Necessity of Virtue: Aristotle and Kant on Virtue*, Cambridge, Cambridge University Press.
Slote, M. (1997), "Agent-Based Virtue Ethics", in Crisp, R. and Slote, M. (eds.), *Virtue Ethics,* Oxford, Oxford University Press, pp. 239-241.
Watson, G. (1990), "On the Primacy of Character", in Flanagan, O. and Oksenberg Rorty, A. (eds.), *Identity, Character, and Morality: Essays in Moral Psychology*, Cambridge, MA, The MIT Press, pp. 449-469.
Williams, B. (1981a), "Persons, Character, and Morality", in *Moral Luck*, Cambridge, Cambridge University Press.

Williams, B. (1981b), "Moral Luck", in *Moral Luck,* Cambridge, Cambridge University Press.
——— (1995), "Acting as the Virtuous Person Acts", in Heinaman, R. (ed.), *Aristotle and Moral Realism,* Boulder, Westview Press.
Wood, A. (1999), *Kant's Ethical Thought*, Cambridge, Cambridge University Press.