

Part I
Basic Themes

COPYRIGHTED MATERIAL

I

Good Will and the Moral Worth of Acting from Duty

Robert N. Johnson

The first section of the *Groundwork* begins “It is impossible to imagine anything at all in the world, or even beyond it, that can be called good without qualification – except a *good will*” (G 4:393). Kant’s explanation and defense of this claim is followed by an explanation and defense of another related claim, that only actions performed out of duty have moral worth. He explains that actions performed out of duty are those done from respect for the moral law, and then culminates the first section with a formulation of that law, “*I ought never to act except in such a way that I could also will that my maxim should become a universal law*”. Kant dubs this fundamental principle of morality “the Categorical Imperative”.

What does Kant mean by a “good will” and what is the meaning and significance of his claim that it is “good without qualification”? And how plausible is the position that only such a will possesses that sort of value? What is “acting out of duty alone” and what is the meaning and significance of the claim that this possesses “moral worth”? And how plausible is Kant’s position that only acting out of duty alone possesses moral worth? Finally, what do these positions and arguments have to do with the stated aim of the *Groundwork*, to articulate and defend the fundamental principle of morality, the Categorical Imperative?

In what follows, I try to answer these questions. My answers will require me to take detours through moral psychology and metaphysics as well as to draw on theories of symbolic expression. My hope is that these efforts will produce a unified overall picture of the main point and purpose of Kant’s discussion of the good will and the moral worth of acting from duty. My plan is this: In the first section, I discuss some initial puzzling claims that Kant makes about the unique value of a good will. I then turn to more basic topics. In section 2 I briefly discuss Kant’s views on the nature of the will, and in section 3 I discuss his views concerning the principles according to which the will operates. In section 4 I take up Kant’s general theory of value and in section 5, the more specific conceptions of unqualified and intrinsic value. I am then ready, in sections 6 and 7, to turn from the topic of the good will to the moral worth of acting from duty. I address

successively two questions here, What is moral worth? and What is acting from duty? In these sections, I speculate about what role these topics have in the *Groundwork's* stated goal, uncovering and defending the fundamental principle of morality. Explaining the role of these topics leads me in section 8 to think about how actions “express” or “exemplify” the principles that motivate them. I end with a summary of how, given all of the above, the structure of the argument should be understood.

1. The Good Will

As a first approximation, to have a “good will” is in some sense to have a strong commitment to behaving morally. So Kant’s claim that a good will alone has unqualified goodness amounts to the claim that only having such a commitment is of unqualified value. He appears to hold that having this strong commitment to behaving morally is what we ordinarily think of as being a morally good person (G 4:402). However, his position has some surprising consequences. For instance, Kant also holds that having a good will – indeed, having “the best will” – is compatible with lacking moral virtue (virtue, as he conceives of it, is the moral strength of will to overcome desires running contrary to what one ought to do) (MM 6:390, 408). So if possessing a good will is what makes one a morally good person, then it seems that Kant embraces the surprising position that one can be morally good, that one’s will can possess “unqualified value”, and yet lack moral virtue. Qualities such as courage and kindness play no part in giving our wills unqualified value and so no part in making us morally good persons. Further, Kant lists other qualities many associate with being a good person, qualities such as “moderation in affects and passions, self-control and calm reflection”, and declares these and like qualities are “conducive to this good will itself” and that they “can make its work much easier” (G 4:393–4). These declarations, however, imply that the possession of a good will can survive their absence (though making its “work” more difficult). Indeed, Kant claims that a good will, because of the weakness of the character in which it is housed, might be utterly unable and unsuited to carry out any of its noble goals, goals such as fostering the wellbeing of those around her and improving herself and her own character (G 4:394). Nevertheless, and again surprisingly, the person possessed of such a will would be a morally good person.

Indeed, a person could apparently lack a good will yet possess the whole panoply of desirable qualities one might naturally associate with being a good person, qualities such as kindness, compassion, courage, moderation, strength of will and so on – the very qualities that Kant himself thinks make one able and suited to pursue the very goals that would be adopted by a good will. Since those are the very qualities of character and temperament that allow one to achieve these noble goals, it is even conceivable that someone might have all of these desirable qualities *and* achieve all of the noble goals a good will *would* have, yet *still lack a good will* and hence *still fail to be a morally good person*.¹ To be sure, Kant thinks that pos-

sessing such qualities and achieving such goals are good things. But Kant asserts that while we might have a favorable view of a being with such qualities, our esteem would be constrained by her lack of a good will (G 4:394). On the assumption that Kant thinks a good will is the lone source of a person's moral goodness, at least to many contemporary readers this is astonishing, since a person with such admirable character traits would strike many as being morally good *just because* she possesses those traits. But according to Kant, those traits apparently have nothing to do with whether she is a morally good person.

All in all, these considerations show that, whatever it is that Kant is thinking of as a good will, it is neither necessary nor sufficient for possessing any other qualities of character that one might reasonably have assumed are intimately connected to moral goodness. But if Kant's views are so far surprising, then it will be even more surprising to learn that Kant believes ordinary folk agree with him. He insists that the views he discusses in the initial pages of the *Groundwork* are nothing but data form from "ordinary moral consciousness", views any reasonable adult would agree with and recognize as reflecting her own views. No doubt, this position may be more plausible with regard to features such as intelligence, creativity or wit. Possession of these is surely ordinarily thought of as being neither sufficient nor necessary for moral goodness, nor are they thought of as having the special, unique value many of us think being a good person has. But it is much harder to believe that we ordinarily think that the traits of being strong-willed, compassionate, courageous, or thoughtful are not important parts of what makes someone a morally good person. We need to know, then, what exactly this good will is, given it does not require possessing the virtues or other desirable character traits, and then to evaluate whether his claim is at all plausible, as a view from ordinary moral consciousness or otherwise. That, in turn, requires us first to understand what it is to possess a will at all. And it is to this topic to which I will now turn.

2. The Will

Unlike events that operate according to natural laws, "only a rational being has the capacity to act *in accordance with the representation of laws*" (G 4:412). By this, Kant means the will is the capacity to choose to act on the basis of policies, plans, or (as Kant himself says) practical principles. Since reason is required to choose on the basis of such representations, such a will belongs only to a "rational being", such as ourselves (G 4:412). The capacity to act on the basis of some plan is different from the capacity simply to choose to act, which Kant thinks that animals also have.² Animals, for instance, do not choose to act on the basis of policies they come up with or because they have reasoned strategies behind their choices (MM 6:213–4). Their choices respond directly to their desires, and these responses, along with the operation of their desires, should in turn be wholly explicable by natural laws: their choices "work in accordance with laws." Human choices, as Kant sees it, never respond to desires in a way that are wholly explained by natural laws, though our desires do affect the formation and execution of the

plans that guide our choices. This distinguishes the human will from a holy will, if there were such a thing. A holy will would make its choices on the bases of policies, as does the human will. However, it would lack any desires that could alter the formation and execution of policies that are rational to follow. Thus, a holy will would *necessarily* make rational choices, lacking as it does any basis for straying from such choices. The human will is thus different from both the non-rational power animals have to make choices directed immediately by desires, and the power a holy will would possess that would be necessarily rational. It is a power to act through choices on the basis of reasons *while affected, but not determined*, by non-rational desires.

Kant sometimes suggests (or at least many have read him this way) that his claim that our wills operate unlike everything else in nature is primarily a metaphysical claim, the claim that the will stands outside of the natural world but nevertheless affects it. In Kant's overall theory, there is a world insofar as it is an object of possible experience, and there is a world as it is "in itself", apart from its being an object of possible experience. And as with everything else that is observable, there is a way that the will is in itself, apart from how it appears to us in our observations of the actions agents (including our own actions). The puzzling suggestion is that this will as it is in itself nevertheless bears some sort of unique relationship to the operations of the observable world, a relationship that will enable us to attribute observable events to the unobservable will as it is, in itself. If this is Kant's position, it is certainly difficult to make sense of, much less accept. It is difficult to even imagine a relationship between observable events in space and time – the operations of an observable act of choice and action – and something that does not exist in space and time and cannot be characterized as being an "event" or an "operation."

However, suppose we were to reject Kant's metaphysical claims. It does not follow that we must reject the entirety of his views concerning the uniqueness of a rational will. One critical aspect of those views is that, as rational beings, we are related to the world in two distinct ways, as observers of the world as well as agents who change the world. We are beings who must think about not only what there is in the world and our place in it, but also about what to do with, to, or about that world. And discovering everything there is to discover about the world and our place in it will not answer the question how we are to change it. There is nothing that faces this question except a rational being with a will, and that makes it unique. We choose what to do guided by principles about what we ought to choose, not simply by what natural laws say we do or will choose. Kant describes this situation as our possessing wills that are not bound by natural laws. But we need not believe that we really *are* such a being to justifiably reason about how to affect events around us. At any rate, in attributing unqualified value to the good will, I assume that we are *not* attributing an extraordinary property to an unknowable capacity that operates only in an unseen, non-natural world. The entire discussion of the good will and its value – as well as the moral worth of actions from duty – is simply addressed to those facing the question of what to do, rational agents.

Kant does not make it clear in his discussion whether a good will is contained in a single choice, whether it consists of a collection of choices over time, or in a disposition to make choices of a certain kind. If he is talking about a single choice, he would be claiming that the only things with moral value are certain particular choices human beings make at particular times. But it is hard to believe that he thinks these isolated events are of such momentous value. It is hard to believe because it might be a fluke that a person makes a given choice for a given reason. Although we cannot rule out that this is what he is talking about, it would be more plausible to suppose he is talking about the collection of choices a person makes over some sufficiently long period of time. When we think of whether a given person has a good will, it seems as if we should be thinking about a whole set of choices, not just one, even if her good will is revealed in each choice. Yet even a set of choices is not the deepest fact about a person's agency. You may not really know the nature of a person's will even though you know what they have chosen and why over a long period of time. After all, she may not have faced any difficult circumstances during that stretch of time. So it seems to me that Kant most likely is talking about a *disposition* to make choices of a certain kind, a disposition to act on certain policies and not others, to pursue certain kinds of goals and not others, and so on. By "disposition" I mean whatever feature it is of a person that makes it true that, if she were in a range of circumstances, she would choose what to do on the basis of a certain sort of reason. Thus, even if she never had actually faced difficult circumstances, for instance, it would still be true of her that had she faced such circumstances she would have made her choices on the basis of the right principles. A good will is thus the disposition to choose courses of action on the basis of certain sorts of policies. This is most likely the thing that Kant believes has such dramatic value.

Kant is aware, however, that his readers could easily misunderstand his claim that this disposition is without qualification good. So early on he points out that he does *not* mean that human beings are necessarily better off for having this distinctive capacity. Indeed, he thinks we would be better off if nature had implanted in us instead an instinct to pursue our own happiness (G 4:395–6). Looking at human history, it may seem hard to argue with him. In any case, Kant argues that those who might think that the will has a natural purpose, and that this purpose is to make us happy, will have difficulty explaining why we would likely have been better off guided by an instinct whose purpose is to insure our happiness. Whatever the natural purpose of a will might be, it ought to be something the will seems reasonably well designed to produce, and it does not seem so well designed to produce our own happiness, indeed especially given that reason seems so often to counsel us *against* proposed ways of pursuing our happiness.

3. Principles, laws and maxims

I use terms such as "policies" or "plans" when talking about what Kant thinks directs human practical deliberation and choices. I use such terms because I believe

they best connect Kant's concerns with our own way of thinking about moral psychology. But it is worth noting that Kant himself envisaged the nature of practical reason as quite complex. Kant distinguishes two sorts of practical principles. Kant dubs principles insofar as they are valid for some person or finite set of persons *subjective* practical principles or *maxims*. Principles that are valid for *every* rational agent he terms *objective* practical principles, or *laws*. We can understand this distinction in terms of plans: Plans to perform actions insofar as they happen to be my, your or some finite set of persons plans are one sort of plan. My plan to go to dinner and a movie tonight is this sort of plan. Plans that are valid for every rational agent are a different sort of plan. Every rational agent, for instance, plans to take the necessary and available means to her goals. We could say that the latter is in this sense an *objective* plan while the former is a *subjective* plan.

For instance, my plan to return a wallet to its owner might be oriented around my desire for a reward. Spelled out as a practical principle or maxim, that would be something such as "I return property when it will likely benefit me." It is the sort of thing a minimally rational agent might well say about returning a wallet, when truthful, articulate and self-aware, in response to a request for an explanation or justification of her action. This would obviously be a plan many other persons would have as well. But it is not a plan that everyone sees as something they have to adopt and act on insofar as they are rational. Some find a plan to be courteous more compelling, others plan to be honorable; neither need have any interest in a reward. Indeed, for the wealthy, a reward will likely not figure in any plans to return a wallet.

Now, contrast the above plans with the plan to return property to its owner (assuming an ordinary context where the person is not a fugitive from the law or some such thing). This would be a plan that is valid for anyone, to return a wallet to its owner, or at least Kant would argue, *even for those who have no interest that would be served by returning property to owners*. If it is a valid plan for anyone to return a wallet – that is, if there are any plans that apply to any rational agent in the circumstances – then we could say it is an objective plan. In the language of principles, we would say that the principle of returning things to their owners, which applies to anyone, is a *practical law*, while the principle "I do whatever will bring a reward," which applies to some, is only a maxim.

Sometimes, the policy some person is acting on in some circumstance could also be *anyone's* policy, as it would have been had my policy been to return property to its owner. In that case, I would have made my choice on the basis of what Kant refers to as a practical law, or at least my policy *could* be such a law.³ Since the will in human beings is the power to choose on the basis of principles, to say that a will is *good* is to say that a disposition to exercise this power of choosing in a certain way is good. It is a disposition to adopt and act on the right sorts of policies. And the right sorts for Kant are or could be practical laws, policies that could or would be those of any rational agent. We shall return to these considerations when we take up the topic of actions from duty.

A *good* will, then, is a disposition to choose that is good because it is based on a principle that is or could serve as a principle for anyone. In order to determine

the nature of the principles of a good will, we will need to understand better the nature of the extraordinary value that Kant is attributing to the good will. Kant begins the *Groundwork* by discussing the peculiar nature of the value of a good will in order to help us to locate what it is about the good will that is the source of this unique value. Thus, by understanding the nature of this value, we may be able to locate the nature of the principle that makes a good will good.

4. Kant's Conception of "Good"

Before I discuss the claim that a good will is unqualifiedly and intrinsically good, I need to explain Kant's conception of what goodness itself is. It will be of little use to try to understand what "qualified", "unqualified", "intrinsic" and "extrinsic" goodness are without first understanding what this property "goodness" is that these adjectives modify. In the *Groundwork* early on Kant appears to connect value with ideal approval. Happiness cannot be good without qualification because "an impartial rational spectator can take no delight in seeing the uninterrupted prosperity of a being graced with no feature of a pure and good will" (G 4:393). The idea would be that a thing is good if and only if an impartial rational spectator takes delight in it. But it is certainly not obvious that this is the analysis of goodness he is offering, especially since he does not explicitly offer it as an analysis. It could just be that good things are things in which impartial rational spectators take delight, but that is not what makes them good.

Later on, he does make substantive claims about value that look like a proffered analysis. For instance, he states that

practical good . . . is that which determines the will by means of representations of reason, hence not by subjective causes but objectively, that is, from grounds that are valid for every rational being as such. It is distinguished from the *agreeable*, as that which influences the will only by means of feeling from merely subjective causes, which hold only for the sense of this or that one, and not as a principle of reason, which holds for everyone. (G 4:414)

Although one could read this as offering an analysis of the property of goodness itself, as I will explain in a moment, I believe Kant is speaking here, not of the property of goodness, but of the properties of the things that possess goodness, that is, the properties in virtue of which, or because of which, a given thing is good.⁴ A good route to the airport, for instance, is a route that, other things equal, has the property of being quicker than other routes. It is in virtue of the property of being a quick route that the route is a good one. A good material for flooring is one that, other things equal, has the property of being durable. It is in virtue of the property of being durable that a flooring material is good. If, as I think, Kant is in this passage talking about such properties, then he is saying that the properties of something that make it good are those properties that determine the will by means of representations of reason.

By “representations of reason”, Kant is referring to practical principles. Hence, the properties of something that make it good are, in his view, those properties that determine the will by means of being incorporated into practical principles. For instance, if I adopt a practical principle of taking I-70 to the St. Louis airport when traveling on the grounds that it’s the quickest route, the property of being the quickest route is what determines my decision by being incorporated into that principle. By “determines the will”, he means “provides a sufficient consideration for the will, insofar as it is rational, to issue in a volition”. Thus, the properties of something that make it possess the property of goodness are, in his view, those properties that provide sufficient considerations for willing by means of practical principles or maxims. Finally, by “practical” good, Kant is here just distinguishing the account of moral and prudential goodness from other sorts of value, for instance, aesthetic value (as in a “good painting” or a “good symphony”), properties that are not sufficient consideration for actions, but perhaps considerations in favor having some attitude toward an object. That which is good (in this practical sense), then, is that thing which has properties that, through being incorporated into our practical principles, provide sufficient considerations in favor of willing it. Kant reasons that, because the properties of good things provide *objective* considerations in favor of willing, those properties provide considerations that are “valid for every rational being as such” or are “universally valid” for rational beings. Objectivity in practical matters at least is, for Kant, thus a matter of validity for every rational being in virtue of their shared rationality.

Thus, so far we do not have an analysis of “good”, but we do know what sorts of properties a thing must have to possess goodness. The *Critique of Practical Reason*, however, contains an extended discussion of Kant’s conception of the good, and helps to fill out the view. Here he states that by “the good”, “one understands a necessary object of the faculty of desire [*Begehrungsvermögen*] . . . according to a principle of reason” (CPrR 5:58). Simplifying this a bit, we can say that the property of goodness is the property of being the necessary object of a rational will. An “object of the will” is simply whatever it is one wills – most immediately, an action. Thus, Kant states,

good or evil is, strictly speaking, referred to actions, not to the person’s state of feeling, and if anything is to be good or evil absolutely (and in every respect and without any further condition), or is to be held to be such, it would be only the way of acting, the maxim of the will, and consequently the acting person himself as a good or evil human being. (CPrR 5:60)

So to be good in the sense in which the good will is good is to be the object of a rational will. Moreover, it is to be the *necessary* object of a rational will. By this, Kant means that it is what a rational will *necessarily* wills to do. Because what is good is what is *necessary* for a rational will, it is what all rational wills, insofar as they are rational, will. Putting this together with the above account of the nature of properties that make a thing good, we get the view that goodness is the property of being, in virtue of possessing other properties that provide universally valid

considerations sufficient for willing, the necessary object of a rational will. At the moment, this may seem quite abstract. However, put together with the rest of Kant's views, its importance for understanding those views will become clearer as we go along.

5. Kant's Conceptions of Unqualified and Intrinsic Goodness

Let us, then, return to Kant's first statements about the unique value of a good will. There is a range of things that we think of as being "good and desirable in many respects", he says, but which can also be "bad and harmful" (G 4:393). "Gifts of nature" such as intelligence and decisiveness, for instance, are not good when attached to a terrorist. Even "happiness" or "total wellbeing and contentment with one's condition" can make a person "bold but consequently often reckless as well" or might be found in a "creature" that never feels "the slightest pull of a pure and good will". In these cases, Kant concludes, happiness is not good. In the former case, it seems Kant thinks happiness can be its own worst enemy, leading the person who enjoys it to recklessness that will undo it. In the latter case, a happy person who does not fall prey to the dangers of bliss still might not deserve their happiness. That a war criminal lives out his days undisturbed somewhere in South America is a terrible state of affairs, but it is made even worse, not better, if he is also happy.

It will be useful to compare Kant's position, that good things such as decisiveness or intelligence become bad when combined with other characteristics, with the doctrine of organic unities, offered by the contemporary philosopher G. E. Moore. Moore's doctrine holds that the value of a given whole is not necessarily equivalent to the value of the sum of the values of each of its parts.⁵ Thus, adding a good thing, such as pleasure, to a bad overall situation, such as a terrorist, does not necessarily make things better. Indeed, many have the strong intuition that adding a good thing makes the resulting whole even worse than the prior situation. So far, this is consistent with Kant's view. However, there are a number of things we might want to say about what's going on in such cases. We might say that pleasure is good but loses that value when it is present in someone wicked, like a terrorist. Then the goodness of it, the pleasure, evaporates and pleasure comes to possess a new property, badness or perhaps neutrality. By contrast, we might say something quite different, namely that pleasure in fact retains its goodness, even in a terrorist, but the resulting whole is not improved or is made worse by adding this good thing. I believe that Kant's position is the former and G.E. Moore's position is the latter. If that's right, then Kant's view has the advantage of being less paradoxical than Moore's. To be sure, Moore's view in a sense explains why the pleased terrorist is a worse whole than the neutral or displeased terrorist: a bad person is in possession of what is in fact a good thing.⁶ It is precisely this that makes the former worse than the latter. However, what is paradoxical is that, if adding pleasure made the situation worse, surely the pleasure itself is what

made the difference. So then it must not be good in this situation, but bad. How else could things get worse by adding it?

Whichever view of what is going on is correct, Kant's view appears to be that a thing's value is qualified just in case one could imagine some circumstance in which it is not good. If this is impossible to imagine, then it is good without qualification. Assume that something is possible when and only when there is some circumstance – actual or hypothetical – in which it can be conceived to exist or occur. It seems when Kant denies that one can imagine a good will failing to possess value, he is supposing this as a test of whether this is possible. Thus, if something is good without qualification, it is impossible for it to fail to be good; it is necessarily good. Kant's claim seems to be, first, that there is such a thing as being necessarily good, and second, that the only thing that has this property is the good will.

The very idea that anything could meet this standard should and no doubt will arouse skepticism. How could anything be good *no matter in what circumstances* in which we imagine it to exist? Surely we can imagine circumstances in which even a disposition to make choices based on objective considerations would not be good (indeed, no matter how we cash out the idea of a good will, it seems we should be able to do this). Imagine, for instance, that an evil all-powerful demon will cause eternal pain and suffering for everyone on the planet if you retain your good will. This disposition seems then not to be good for you to have. Outlandish as this fanciful idea is, it is *conceivable*. So it seems there is a possible circumstance in which being morally upright is not at all a good thing. For this reason, it seems likely that there will be conceivable circumstances in which *anything* is not good. So if this is what Kant's notion of "good without qualification" comes to, it appears to be in trouble from the outset.

Perhaps, then, Kant means something else. Perhaps he wants us to distinguish the valuable quality of having a good will itself – for argument's sake, the quality of being disposed to act on principles that are objectively valid – from its being possessed by someone – a particular person's actually having this disposition. Thus, while the quality of possessing this disposition in itself might be valuable no matter the circumstances, some particular person's actually *having* that quality is in some cases not valuable. This would allow us to claim that, in the demon case, it is that particular person's *possessing* a genuinely valuable quality – that of being morally upright – that is bad in the circumstance, not the quality itself. So Kant's claim would come to the claim that while the *quality* of being morally upright itself is good in any conceivable circumstance, *possessing* that quality would not be good under any conceivable circumstance.

What could it mean to say that some quality is good though not possessed by anything? Perhaps it means that it is "good in the abstract." Thus, "having a good will" is good in the abstract while "failing to have a good will" is bad in the abstract. Even if this makes sense, the problem would be that we seem bound to say the same thing of intelligence and all of the other qualities Kant cites as having only qualified value. For instance, we should not say that intelligence or pleasure is often a good thing, but not when it is in a criminal. Rather, we should instead

say that though it is good in the abstract, it is not good for a criminal *to possess* it. And so on and so forth through all of the things Kant lists as qualified goods. That leaves us with no way to distinguish between qualified and unqualified values.

It may be that Kant means that while we can at least conceive of being intelligent, being happy, or any other putative good property as *in itself* failing to be good we cannot conceive of the property of having a good will as *in itself* failing to be good. The idea would rely on a general fact about many qualities: whether a thing possesses those qualities depends on what is going on around them. For instance, a table possesses the property “being next to the sofa” depending on what’s going on around it – that is, where the sofa is relative to it. Put the table in front or in back or on the side of the sofa and it has the property; put it on the curb and it no longer has that property. Some qualities may change from being good to failing to be a good in the same way. Thus, the property of being good may change in most things depending on what’s going on around the feature (such as intelligence) that is good. The claim, then, would be that unlike every other thing that is good, the property of goodness in a good will remains no matter what else is happening around the person who has it. In particular, it remains even if there is a horrific outcome – as, for instance, in the imagined demon case. We say that it is a bad thing that the person in the demon case possesses this good quality, rather than that the quality itself is no longer good in this circumstance. And this would then distinguish the good will from things such as intelligence and pleasure. It would do so because there are cases in which it is not only or merely that the person’s possessing intelligence, and so on, is bad, but in which the intelligence itself would fail to be a good thing.

One might say this, for instance, of events that take place in the romantic novel *Frankenstein* (or at least it is one way of thinking about Mary Shelley’s work of that title). Dr. Frankenstein’s vast knowledge of what animates living beings loses its value because it is had in an unnatural degree. It is a good thing in quantities that fall short of that which Dr. Frankenstein had, perhaps. Somehow, knowledge can lose its value when it concerns altering the order of nature. Then, it is not merely that it is a bad thing that Frankenstein possesses this valuable thing, knowledge; knowledge itself has lost its value, and even has acquired a disvalue. That, at any rate, may be the way we are supposed to understand Kant’s idea that the value of a good will is unique. There is a special sort of value that the good will has such that this value is not based on the circumstances in which the good will is placed. The view isn’t simply that the good will has the same sort of value as other good things, and only differs from these other things in possessing that selfsame sort of value in every circumstance. It is instead that it possesses a value of an entirely different order, a sort of value the possession of which doesn’t vary with circumstances. This would be a significant feature of the goodness of the good will, if the good will possesses it. But it is still unclear what the nature of this sort of value might be.

Kant also refers to unqualified goodness as *unconditional* goodness. Perhaps this idea will help us to understand his view better. The idea is not unfamiliar.

Something is conditionally good just in case there is some condition that must be met for it to be good. Its value depends on a condition, typically on the condition that some other thing is good. So a given good thing is conditionally good when some other thing's goodness is a condition of the given good thing's goodness. For instance, surgery is a good thing – but only on the condition that some other thing, the thing it produces, is good, such as health (assuming there is no such thing as recreational surgery). Money is a good thing – but, again, only on the condition that there are other valuable things, things that it can be exchanged for. Otherwise, it is just paper and metal, as it might become were one stranded on a deserted island with a fat wallet. A daub of color in a painting might be good, but only on the condition that some other thing, the painting of which it is a part, is good. In each case, there is some good thing – health, things to buy, a painting – whose value is the condition of the value of some other good thing – surgery, money, a daub of color. Kant claims that, in some such way, everything other than a good will has a condition of its goodness. The will's goodness, additionally, is supposed to be the condition of the goodness of *everything* that has value. That is, Kant believes that nothing is good unless the will of the person who possesses it is good. And, unlike every other valuable thing, there is no further condition of a good will's value. That is, there is no other thing distinct from the good will such that its value is the condition of the good will's value.

That a thing of qualified value has a condition under which it is valuable would explain why it is possible for it to lose its value, why it is only good in some circumstances but not in others. It is possible for it to lose its value because it is possible for the condition of that value to be missing in some circumstance. It is possible for surgery to be bad, for instance, because we can imagine the surgery without the circumstance of producing a benefit to our health. Think of the myriad of malpractice claims on this basis. That circumstance is the condition of surgery's having value. But the idea of being valuable only in some circumstance can still be distinguished from the idea of being only conditionally valuable. It can still be distinguished because the intuitive idea of a circumstance is the entire state of affairs surrounding a thing, but the condition of a thing's goodness seems to be only some constituent of that whole state of affairs, the particular constituent of the state of affairs upon whose value that thing's value depends.

Kant also claims that an unqualified good is good *in all respects*. That suggests that if a thing is of qualified value, then there is something about it that is not good, some respect in which it is perhaps even bad. The thought experiment of imagining a thing in a different circumstance could be thought of as a way of revealing respects in which it is not good.⁷ Perhaps this is what Kant is thinking about the good will. We can discover, by doing these thought experiments, that there is nothing but a good will that is *all* good, so to speak. For instance, knowledge is supposed to be only of qualified value. When we imagine it in a Josef Mengele or Dr. Frankenstein, for instance, we think that intelligence isn't all good. There is some facet or respect in which it is not good. Now the respect in which knowledge is not good may not be apparent in the circumstances in which we judge it to be of value. But all that we have to do is to imagine it in different cir-

cumstances, and the respect in which it is bad will be plain. Shelly imagined forbidden knowledge, for instance, aberrant and unnatural. Perhaps she thought that at a certain point, knowledge can separate us too much from the rest of the natural world and that is a bad thing. Kant himself thought that intelligence can be “extremely bad and harmful when the will which makes use of [it] . . . is not good” (G 4:393). Part of what he is pointing to is a purely extrinsic feature of intelligence, its usefulness to pursuing purposes, good or evil. But as the fanciful case of the all-powerful demon indicates, there are extrinsic features of a good will that might be bad in this way too.

However, this now suggests that we have discovered an important difference between a good will and every other valuable thing: The person who possesses intelligence *can* use it in order to pursue his own evil purposes. There is nothing about the very nature of knowledge that bars its use by the wicked. But while some other person might make use of a person’s good will for evil purposes – the demon, for instance – it is hard to imagine how the person who possesses a good will herself could ever, in any circumstance, use her good will for evil purposes. It seems that its very nature rules out this possibility, while the very nature of knowledge does not. This would make a good will unique – if, that is, every other thing or trait a person might possess could be used to pursue some evil end. If being good in every respect is at least part of the notion of unqualified goodness, then Kant’s claim is that the good will alone is good *in every respect*. There is no respect in which it can be put to work in the name of something wicked by the person who possesses it. It seems, then, that for something to be unqualifiedly good no condition of its goodness should be lacking and it must be good in all respects.

Another aspect of Kant’s views on the value of a good will is that it is *intrinsically* good, in the sense that it is “good in itself” (G 4:394). Is this the same thing as being unqualifiedly good? Kant’s discussion gives the impression that he thinks that these concepts are different. He deliberately takes additional space to explain why an unqualifiedly good will is good in itself. Kant clarifies what he means in saying that a good will is good in itself by saying that it is good “only by virtue of its willing” (G 4:394). Being disposed to volition seems to be an intrinsic property of the will. So at least part of his meaning in saying that a good will is “good in itself” and “only by virtue of its willing” is that a will is intrinsically valuable when its intrinsic properties are what make it valuable.

A standard conception of an intrinsic property is of a property of the sort we have already characterized as the goodness of an unqualifiedly good will: A thing’s intrinsic properties are those properties that it could still have regardless of how we might change the circumstances around the spatio-temporal region inhabited by the thing. For instance, no matter whether a baseball is in a glove, on the ground, in the air or in the stands, it is still spherical. Its sphericity does not change when the circumstances around the spatio-temporal region inhabited by the baseball change. But a given baseball would lose the property of “being inside of Busch stadium” once someone hits it out of the park. Its “being inside of Busch stadium” changes when the circumstances around the baseball change. So that property is an extrinsic property of a baseball. The property of being white is an

intrinsic property of the baseball, since it is white wherever it is, but being smaller than a basketball is an extrinsic property, since a baseball would lack that property were there no basketballs. Intrinsic value, as Kant means us to understand it, seems to be the value a thing has in virtue of properties it could retain no matter how its surroundings might be changed. And the only intrinsic property of a will is the volition that characterizes it.

A caveat. Note that this understanding of intrinsic value does not yet require the property of value itself to be an intrinsic property. It is consistent with everything I have said so far that the property of value itself is in fact an extrinsic property of a thing that has it. For, in general, a thing can possess extrinsic properties in virtue of its intrinsic properties. For instance, a lamp may be expensive because of its shape. Or a shirt may be desired because of its color. Shape and color are paradigms of intrinsic properties. Expense and being desired are extrinsic properties. To be expensive is to cost a lot of money, but prices change based on market forces, supply and demand, and so on. A thing is desired if someone desires it. In the absence of someone's desiring it, the thing loses the property of being desired. In fact, intrinsic value might turn out to be like being desired: Things can be desired because of their intrinsic properties; in such a case, we think of them as intrinsically desired. The same might be true of intrinsic value for all that we have so far said about Kant's own conception of intrinsic value.

Now, the properties that make a thing valuable are conditions of its value. But they may not be the only conditions. Suppose a thing's being intrinsically good means that when it is good, it has intrinsic properties that make it good. Then those conditions of the thing's value would be present no matter how circumstances might shift around it. Nevertheless, there might be other conditions of that thing's goodness as well, conditions under which the good-making intrinsic properties of a thing will in fact make that thing good. For instance, what make pleasure valuable are its intrinsic properties. But it might also have some condition under which those intrinsic properties do not succeed in so doing. For instance, the intrinsic properties of pleasure that are the source of its goodness when it is in fact good might not make it good when a wicked person enjoys it. Even in the circumstance in which a wicked person enjoys pleasure, it retains the intrinsic properties that are the source of its value. There is just this further condition under which it is good, namely, that a wicked person is not enjoying it.

In any case, one thing seems clear: something that is valuable without qualification must also be intrinsically good. This just seems to follow from what we've learned about unqualified goodness. For an unqualifiedly valuable thing is valuable in any circumstance, and that means there is no circumstance in which the conditions of its value are lacking. But what remains across changes in the circumstances in which a thing exists will have to be that thing's intrinsic properties – properties that do not change across circumstances. And if that is so then whatever conditions there are of an unqualifiedly valuable thing's value, they must be that thing's intrinsic properties. So an unqualifiedly good thing must be good in itself. We've just seen, however, that the reverse is not true: that something is in itself good is consistent with its not being good without qualification. For a thing, for instance

pleasure, can be good when it is good because of its intrinsic properties, yet have conditions under which those intrinsic properties fail to make it good.

There is a further reason why an intrinsically good thing need not be unqualifiedly good. This is that a thing's intrinsic properties are different from its *essential* properties, or properties a thing *must* have in order to be what it is. The sphericality of a baseball does not change as the circumstances around it change; but a baseball need not be spherical; the game might have evolved with oval balls or some other shape. A baseball is white, but it could have been brown or gold. A thing can lose its intrinsic properties – properties that do not change when the circumstances around it changes – and remain what it is. But if an unqualifiedly good thing is good no matter what the circumstances it is in then it must be not only intrinsic properties in virtue of which it is good, but also those properties must be essential properties as well, properties that must be present in order for the unqualifiedly good thing to be present.

This is a desirable outcome for a number of reasons. First, it seems that there are many things that are intrinsically good – valuable things about which it is true that what makes them valuable are their intrinsic properties – yet there are respects in which they are nevertheless not good. We've already noticed that pleasure appears to have this property. But many other things seem to have it as well. Beauty and knowledge for instance are good in themselves – that is, what makes them valuable are their intrinsic properties. But we do not think they are therefore unqualifiedly good, for they possess features that can, for instance, lead us to violate moral principles. They can be used by the person who possesses them to do evil things.

Second, most things appear to be, when good, in a number of different ways and also in some possible circumstances bad. And the circumstances in which things are good or bad appear to differ from each other in that some of these circumstances involve the intrinsic properties of the thing and some involve its extrinsic properties. The respects in which a diamond is valuable, for instance, seem to be mainly properties that do not change when the circumstances around the diamond changes – its beauty, for instance. But a diamond also is very hard and can be used to cut softer materials such as metal or glass. That hardness is a source of the diamond's extrinsic value, a value it would have only because it is related to some other thing of value – whatever might be achieved by cutting metal or glass. Indeed, it is the fact that most goods have such a mixture that is the source of the fact that they are not all good, not good in every respect.

With these details about unconditional and intrinsic value in place, let us consider again Kant's claim about the good will in full:

A good will is not good because of its effects or accomplishments, and not by virtue of its adequacy to attain any proposed end: it is good only by virtue of its willing – that is, it is good in itself. (G 4:394)

Read literally, this appears to say that not only is a good will intrinsically good; it is *only* intrinsically good – *only* good by virtue of its volition or willing – and

has no other value at all. But how can this be? Surely a good will can *also* be useful, indeed, might also be the best way of attaining some proposed valuable end. This seems odd, especially when we cast this, as I have, in terms of choice. Whether some choice you made was a good choice seems to depend on the value of the thing you chose as much or more than any intrinsic features of your choice itself. And surely no matter how good a choice is it might also be advantageous and hence good because of that as well.

However, here we have been talking about the value of an *unqualifiedly* good choice. A choice could not be good without qualification in virtue of the value of the outcome or any other extrinsic feature of that choice. This is because, for any choice and any outcome or other extrinsic feature of that choice, there will be changes in the circumstances that will not include that outcome or that extrinsic feature, but might still include the choice. Hence, as I argued above, if a thing is unqualifiedly good, then it is also intrinsically good. One thus need go no further than the intrinsic features of an unqualifiedly good will to find that in virtue of which it possesses this sort of value. This is why Kant says that a good will is good *only* by virtue of its willing; its willing is that intrinsic feature of a good will that makes a good will possess unqualified value.

As we've already seen, in Kant's view the will is our capacity as rational beings to perform actions by choosing them on the basis of principles or the "representations of laws" (or universally valid or acceptable plans). One way in which we exercise this capacity is when we choose on the basis of our representation of natural causal laws. For instance, suppose we choose to steer clear of the campfire in front of us because it will burn us otherwise. In that case, we have chosen on the basis of our representation of natural laws such as that fire burns human skin, and that by staying at a distance we avoid its burning our own. Indeed, were someone to ask why we were steering clear, we might well give as a reason "It will burn my arm otherwise." Here, we are making our choice (to steer clear, in this case) into a natural causal link to complete a causal chain that results in a desired outcome (avoiding burns). We are operating according to our representation of a certain chain of causes governed by natural laws. We represent those causal chains because we have a naturally occurring interest or desire for something (such as avoiding pain). We then conceive of that desired thing as something to be caused by our choices, and so conceive of our choices as causes of that thing's coming to be. This is practical reasoning about how to bring about goals that we have set for ourselves, and this is how it operates on the basis of principles.

Kant says that insofar as choices are based on such principles of reasoning, they are "in some way good" (G 4:414). By this, Kant means that they are good for achieving the goals we set ourselves to achieve. "That was a good choice," we might say about avoiding the campfire, for instance. We would mean it was a good choice because it was based on a good principle – to avoid campfires so that we don't get burned. Probably in some sense we might also think of a choice that just happened to result in a desirable outcome as a "good choice". But we wouldn't think it had genuine value as a choice. It would be good in the sense of

being a lucky choice (although many, of course, would rather be lucky than good). A choice is in fact good only when it is good because the principle on which it was based is a good one. And what makes the principle in this case a good one is that it provides guidance to achieving a desirable goal, avoiding burns. Because this is what makes it a good principle, we would not, in saying that the choice based on it was good, mean that it was good in every sense. After all, there are things for which it is worth risking burns.

This is how we bring to bear representations of laws on our goals and thereby arrive at principles that apply to our situation and goals. And we also see the connection to the value of choices: it is because the choice is based on a good principle that we find the choice good. The same, moreover, is true of a good disposition to choose. A disposition toward a certain choice is good because of the goodness of the principle on which the choice would be based. But the good choice and disposition that we're interested in are not just good in any sense, but in an unqualified sense. Given that, they cannot be good because they are based on a principle that is good for achieving some desired results. This is because the condition under which such a choice will actually be good is the condition consisting of being in circumstances in which that choice leads to those desired results. But for any result we might want to achieve, the conditions will vary for achieving them. We may want to avoid getting burned by the campfire, so we avoid it. But it might be that avoiding the campfire requires walking through a hail of bullets. In that case, a choice on the basis of the principle of avoiding fires is not a good one.

What practical principle, then, makes a choice good, not because of its usefulness in achieving our goals, but no matter what the circumstances, and so makes the choice unconditionally good? Kant's answers this question in two steps:

1. "The pre-eminent good which we call 'moral' consists therefore in nothing but *the idea of the law* in itself . . . so far as that idea, and not an expected result, is the determining ground of the will" (G 4:401).
2. The "law . . . the idea of which must determine the will . . . if that will is to be called good absolutely and without qualification . . . is the [law of] universal conformity of actions to law as such" (G 4:402).

Change the circumstances around a given choice based on a principle such as "This will bring me more pleasure" or "That will make others happier", and that choice no longer is effective. So effectiveness in bringing about our goals cannot be essential to that choice. All you need to do is to imagine the circumstances are such that you no longer aim at pleasure or making others happy. Then, a choice based on that principle would provide you pleasure and make others happier, but wouldn't be good. Indeed, the value of choices based on all principles of instrumental reasoning depends on a property such as effectiveness, a property that is not essential to those choices. And that means that whatever the principle is that makes a good will good must be a non-instrumental principle. It must be a principle that makes a choice based on it unconditionally good, a principle such that

no matter how circumstances are changed around the choice based on it, it still retains that property in virtue of which it is a good choice.

The claim that a good will is good without qualification thus at a first approximation means that under every conceivable circumstance in which it exists, a choice characterized by some principle, call it “M” (whatever M turns out to be), is good, and is good because it is characterized by M, and any other choice is good only if it is consistent with choices characterized by M. So, if there is a good will, then there is some feature of a motivating principle M such that it makes a choice based on it a good choice, and it would possess this feature under every circumstance.

Let me briefly summarize Kant’s views as I have so far presented them. An unqualifiedly good choice is a choice based on some principle that makes that choice good under any circumstance. That it is good under any circumstance tells us to look for some property of the principle on which the choice is based that the principle would retain no matter how circumstances might be changed around it. It tells us to look for an essential property of that principle. We thus arrived at the question: What sort of motivating principle has a feature that makes a choice based on it good and is essential to that principle? We already know that it cannot be an instrumental principle, a principle whose value depends on achieving or being aimed at achieving certain results. The value of a choice based on that sort of principle is entirely dependent on the circumstances, and that means the property in virtue of which a choice based on it is good is effectiveness – an inessential property. So only a principle that is a non-instrumental principle makes a choice based on it good because of a property that is essential to that principle. To say more about that principle, we need to turn to the topic of the moral worth of acting from duty.

6. Moral Worth and Duty

Kant holds that if it is good that you do some action, you can only think of this as what you *ought* to do, what you *must* do, or as he often puts it, as your moral *duty*. Kant believes human beings inevitably think in these terms because whenever some action would be good to do, we nevertheless are aware of the fact that we might not do it. We are rational agents, but also creatures with desires, and these desires are variable and contingent. I have desires you do not have and you have desires I do not have, both of our sets of desires change over time, and there is no desire everyone has had and no desire anyone always has. As a result, we have the potential to be motivated by desires that not everyone could be motivated by and fail to be motivated by desires others are motivated by, or we were once motivated by but are not now motivated by. So the thought of what it is good to do can only represent a constraint on what we will do. It is a constraint on what we want for ourselves or for others. The good, as a result, can only motivate us through the thought of duty. This is why Kant says that the concept of duty *contains* the idea of a good will under certain subjective limitations and hindrances (G 4:397). A good will *in us* is a will that might not do what is good,

and so can be only motivated by duty rather than the good. This fact, the fact that the investigation of what motivates a good will in human beings must turn to a consideration of acts performed out of duty, has a side benefit: Kant believes that cases in which a person acts out of duty clearly exemplify the motivating principle behind the action. That in turn makes it easier to discover what the principle is that is the source of the extraordinary value of a good will.

Kant gives us four examples of persons acting in various ways to illustrate the point he wants to make. The first, a shopkeeper, gives correct change to a customer. The second preserves his own life. The third, a Samaritan, helps others. A fourth looks after his future happiness. In each case, Kant tells us that the action in question is a moral duty and asks us to imagine it being done for a variety of reasons. He does this because he wants us to compare the moral worth of the action when it is performed for these different reasons. In particular, he wants us to compare the case in which the person acts out of duty with when she does not. In each case, he believes we will see that it is only when the action is done because it is one's duty that it has what he calls "moral worth."

For now, let us set aside what Kant means by "acting from duty." What exactly does he mean by "moral worth," the property possessed only by an action done out of duty? Many readers will assume that "moral worth" is an ellipsis of "*worthy of*" something.⁸ Kant's position would amount to this: only a person who acts from duty alone is praiseworthy, worthy of happiness or perhaps deserving of a special attitude such as esteem or respect. It is worth pointing out that if this is indeed what Kant means then his discussion of the moral worth of actions from duty is peripheral to the discovery and defense of fundamental moral principles, not central, in other words, to the discovery of what makes a good will good. It would not be central because what ordinary people think we are worthy of or deserve as a result of what we do would shed little light on what the principle is that tells us what we ought to do. Kant's claims about which actions have moral worth would only be of marginal interest, of interest because they are controversial pronouncements about how ordinary people think such things as esteem, praise or happiness ought to be doled out. However, I think that this is the wrong way of understanding the topic of moral worth. I'll explain why.

One common way of thinking about all of this is that to attribute moral worth to an action from duty is, for instance in Richard Henson's terms, to issue either a "battle citation" to or a "fitness report" on the agent.⁹ A battle citation is due someone who victoriously overcomes inner opposition and does the right thing; a fitness report tells you that a person is ready to overcome that opposition, should it arise. Both appear to be forms of praiseworthiness. The idea is that an agent is morally praiseworthy in virtue of some duty she performs when and only when in performing her duty, she has overcome significant obstacles or was capable of doing so. But only when she performs her duty out of duty alone is either of these things true. So only when she acts out of duty alone is she praiseworthy.

If a praiseworthy action is an action that *ought* to be praised then this would be a deeply problematic view. For any moral theory, including Kant's, whether one *ought* to praise an action is a substantive moral question, since praising is an

action and so is itself up for moral evaluation. In the case of Kant's theory, whether one ought to praise an action, including an action from duty, should depend on whether the Categorical Imperative tells one to praise in the circumstances, and it is not at all obvious that it will always or even typically do so. Indeed, it is all too easy to think of cases in which *prima facie* we ought to praise actions that were not done from duty or whose status is opaque. It might well show most respect for morality to encourage people to be moral even when we do not know why they did their duty. And it is equally easy to imagine that we ought not to praise an action that was performed from duty. For praising under certain circumstance can be disrespectful, for instance, because in the circumstance it would be ostentatious or unwelcome. After all, a person may well think it was only their duty to act as they did and that doing one's duty is not out of the ordinary for them, while praising them might under certain circumstances appear to imply that they did something out of the ordinary.

Of course, it might be that his position is that only such actions nevertheless *deserve* praise, and perhaps it is the deservingness that is important, not whether we ought to praise. This is certainly more plausible. But it is not consistent with what Kant himself says on the topic, even in these passages in the *Groundwork*. He explicitly and repeatedly asserts that many dutiful actions not from duty, such as those from sympathy and honor, *also* deserve praise and indeed can be morally meritorious.¹⁰ And surely he was right to have said so. It is not contrary to any other element of his views. But it is also not true that Kant's views require that *all* actions from duty deserve praise, at least if we take him seriously when he says, in the *Doctrine of Virtue*, that weakness in overcoming obstacles to doing one's duty is something "childish and weak, which can indeed coexist with the best will" (MM 6:408). A person with such "childish weakness" might often enough have no other reason to act *save* the fact that it is his duty in some circumstances, and have no countervailing desires or interests. It is not at all clear why this childish and weak person who does his duty would deserve praise for doing it from duty in such a circumstance. It is only his duty, after all, and so he might not be deserving of praise for doing it for that reason alone. Nevertheless, it seems as if he could well act solely from duty and so it seems he could well have done something with genuine moral worth. Those who are critical of the position that all and only actions from duty deserve or ought to be praised are right; it is an indefensible position. But it does not appear to be Kant's.

"A good will" Kant wrote, "seems to constitute the indispensable condition even of worthiness to be happy" (G 4:393). This statement suggests an alternative possibility, that moral worth is to be taken as worthiness of a reward of some sort, such as happiness. Possessing a good will is indeed in Kant's view the condition of your worthiness to be happy. Are then all and only those who act from duty worthy of happiness? Is that what "moral worth" means? It would be surprising if Kant held that it was. Moral worth is an attribute of particular acts of will, and, however pure, a particular act of will is surely insufficient to make one worthy of happiness. Our worthiness to be happy more plausibly hangs on the moral character of our overall life. Moral worth seems concerned only with the condition of

the will in discrete actions, or at any rate, that is the strong impression one gets from Kant's discussion.

Perhaps then one is worthy of happiness to just the degree this action contributes to the overall moral condition of our wills. The overall moral worth of our actions as a set may then constitute how deserving of happiness they make us. When a good willed person strengthens her will and increases her virtue, she appears to become by Kant's lights more worthy of happiness. But even if this is so, it does not seem to be what Kant is concerned with in the sorts of judgments that he asks us to consider. Judgments about our worthiness of happiness are apparently for God to make and must not form the basis for our attitudes toward others. Kant does not ask us how much of a boost to his deservingness of happiness the shopkeeper or philanthropist got for acting from duty alone. We are asked to consider only some quality of the will of the person acting, not of what that quality makes him deserving. All in all, Kant's term "moral worth" is not best understood as elliptical for "worthiness" of some reward either, even if possessing a good will constitutes one's worthiness to be happy.

Actions possessing moral worth deserve esteem; Kant makes this plain. Perhaps, then, for an action to have moral worth is just for it to be worthy – in the sense of deserving – of an attitude such as esteem.¹¹ Now if esteeming someone were understood as an action like praising her, then virtually all of the considerations regarding praise and rewards would also argue against taking moral worth to mean worthiness of esteem. The one consideration that would not is that moral worth *does* guarantee that the agent deserves esteem. However, it does not follow that if the agent deserves esteem for her action, her action has moral worth. If esteeming is not an action like praising, but an attitude, then it is a *response* to moral worth rather than what moral worth is. To "deserve esteem" is to be the fit object of the attitude of esteem; or, insofar as one is rational, one responds to such worthy things with esteem.¹² It is the moral worth of the action that makes it a fit object of esteem. So even if the idea of esteem-worthiness is relevant to which actions have moral worth and which do not, "moral worth" is not elliptical for "worthy of esteem." Esteem is the proper attitude toward actions with moral worth, and could serve as a guide to which actions possess it and which actions do not, but there is the additional matter of what makes actions deserving of this attitude.

The most telling consideration against understanding "moral worth" as "worthiness of" something is the absence of any rationale for its place in the first chapter of the *Groundwork*, a chapter in which he is initiating his search for the fundamental principle of morality. What the reader expects in a discussion that begins with the unique value of a good will and ends with a fundamental moral principle of action is not a discussion of who is worthy of what, but of the nature of moral value as it concerns the actions of human beings. If one is looking for the fundamental moral principle that is to guide human action, and one has begun with the assumption that this will be what makes a good will have its special value, then I believe it is most natural to expect a turn to a consideration of this value as it appears in discrete acts of will. In other words, "moral worth" seems best understood as just another name for moral value, in this case, as it is found in particular

acts of will or volitions, as opposed to a general disposition of the will. And moral value may well be a property a will must have in order to qualify for or engender certain kinds of attitudes, such as moral esteem, but it is not equivalent to being so esteemed.

My proposal, then, is this: Kant's primary objective in discussing actions with moral worth is to discover that in virtue of which discrete acts of will, discrete willings if you will, possess moral value. That then raises the question of *why* certain acts of will are morally valuable and others not. That is, it raises the question, Which actions – the conception of which includes the principle characterizing the choice of this action, or the plan executed by the choice – have a value that would remain no matter how the circumstances might be changed around those actions? *These* willings possessing *this* property are the objects of esteem, the response that befits moral value. The principle behind these willings will be the fundamental principle of morality.

The connection between esteem and moral worth has to do with the fact that esteem is a form of respect, respect for *the person* as focused on the operation her will. Now Kant holds that “all respect (*Achtung*) for a person is actually only respect for the law (of righteousness, etc.) that person exemplifies” (G 4:401n). The idea is that the source of the esteem we have for a person whose behavior displays before us the moral law is in fact just our respect for that law itself. The behavior displays or exemplifies that law when that law alone motivates it, and our esteem for such a person is really a response to the moral law itself. The person who exemplifies the moral law is living up to the standard set by the moral law. But our attitude is focused, not on that achievement as an achievement per se, but as an example of a law that draws the respect of rational agents. If esteem is the appropriate response to moral value, and what is of moral value is that which retains its value no matter how circumstances are altered around it, then esteem is the appropriate response to this intrinsic value of the thing exhibited by a good will. And that is the moral law itself.¹³ That law is what our esteem focuses on. A given person's action has moral value, then, when and only when rational agents would esteem his willing of it, and they will do this when and only when willing his action exemplifies the law that rational agents respect. And that in virtue of which it has value is its intrinsic property, its principle or maxim.

7. Acting from Duty

We will take moral worth to be moral value, then, and now turn to the actions that alone supposedly have this value, actions from duty. Now a very common, perhaps even standard, way of reading Kant's discussion of acting from duty focuses on his claim that actions from other motivating principles such as sympathy are only “fortunate . . . to aim at something generally useful and consistent with duty” (G 4:398).¹⁴ The idea behind this standard reading is supposed to be that actions chosen because we have concern for others' welfare or other seemingly desirable motivations have no moral worth because they only produce actions that

conform to what is our duty by accident. Such principles aren't reliable in producing the right action, as determined by some, perhaps ordinary intuitive, standard. On the standard reading, we come to the examples with some conception of duties and of actions conforming to these duties, and Kant is supposed to want us to consider which motive is connected in the right way to that conformity. Thus, Kant tells us that helping others in need is a duty, but that doing so out of heartfelt concern for their plight has no moral worth. Indeed, it is only when we imagine a person who has lost all interest in her fellow human beings, but still helps them because it is her duty to do so, that we have the image of an action with genuine moral worth.

This picture has led some critics to claim that Kant's views have the absurd implication that we must come to dislike doing those actions that are our moral duty, in order that they might be done solely out of duty. There are several mistakes in such a criticism. As many have pointed out, it assumes that acting solely out of duty is incompatible with having sympathy and other emotions.¹⁵ But Kant makes no such claim, and nothing he says requires this. So we can help others from duty and with sympathy, and *mutatis mutandis* with other duties and other feelings, attitudes and dispositions. Second, Kant does not say or imply that there is anything *wrong* with doing one's duty from other motives, nor that we should try to rid ourselves of such motives and replace them with a motive of acting from duty alone. Third, we have to keep firmly fixed before our eyes Kant's overall purpose, which is to discover what the fundamental principle of morality is, the principle that guides and motivates a good will. He picks the examples he does because in his view they clearly exemplify that principle. It is only when every other motivation counsels against doing one's duty that we can see and be sure that the person's sole motive was simply that it was his duty.

Much of this way of understanding Kant's point is right, I think. But this way of defending Kant and the criticisms that provoked those defenses are ultimately both flawed. If Kant's point is that acting from duty is acting from the only reliable motive in producing dutiful actions, then it is, again, difficult to understand what relationship the discussion of the moral worth of acting from duty (under this interpretation) would have to the primary point of the *Groundwork*, the discovery and justification of the fundamental principle of morality. Indeed, the discussion of acting from duty presupposes the existence of some principle that defines our duties. It presupposes this because it asks us to assume that the actions in question conform to such a principle, but are based on motives that are unreliably connected to producing this conformity. Moreover, the discussion, so understood, does not seem to connect to the discussion of the extraordinary intrinsic value of the good will, a value that could not be connected to its reliability in doing anything given that reliability is quite clearly an extrinsic property.

As I read Kant's examples and discussion, he assumes no independent gauge of right and wrong at work in these examples. In my view, Kant's point, and this is important, is simply that the principles motivating the shopkeeper, the philanthropist, and so on, are not principles that underlie an unconditionally and intrinsically good choice. The argument that they are not is based on the fact that in each case

there are conditions on the value of the choice. Kant says that they only fortunately hit on something useful and dutiful because he wants to make the point that there is a condition of the value of such choices. It then follows that the principle on which these choices are based cannot be the principle of a good will. And if the principle on which these choices are based is not the principle of a good will, it also cannot be the fundamental principle of morality. For Kant's guiding assumption is that *whatever it is* that guides the choices of a good will, it will be the fundamental principle of morality, the very principle that he is trying to discover and justify.

Consider, then, an alternative understanding of Kant's discussion. Kant's claim that only actions from duty have moral worth, on this alternative understanding, amounts to the claim that, among the actions of human rational agents, only actions from duty express the unqualified value of a good will. If this is so then it appears that only the actions of someone with a good will possess moral worth. If I am right that a "good will" is an enduring disposition to act on the right principle or plan, and "the right principle" turns out to be "it is my moral duty" then this in turn implies that only someone with an enduring disposition to act because it is her moral duty to do so performs morally worthy actions. When her actions are the outcomes of her disposition to act because it is her moral duty to do so, then they have moral worth because they express the value of her good will.

I myself think that it is not an accident that Kant uses the examples he does to explain what it is to act from duty. Many philosophers, such as Hobbes, think that self-interest is a primary, even unitary, motivating principle behind all of our actions. If that were true then to justify moral principles – to show why we should perform them – would require showing that they are grounded in self-interest. Moral principles are those that ultimately serve your interest, such philosophers argue. The same is true for self-preservation and happiness. Sympathy, too, has been an element of the moral psychologies of some philosophers such as Butler and Hume. Moral principles are in their view somehow grounded in our sympathetic natures. Thus, in my estimation, each of Kant's examples is a challenge to each of his opponent's theories, theories that claim to ground moral principles in the motivation it illustrates. Is the fundamental principle of morality – the principle that explains and justifies our moral duties – at bottom the principle of self-interest? Is it a principle based on self-preservation, on our sympathetic natures, or on our desire for happiness? The discussion of moral worth is a shot across the bow for proponents of any such theories. Kant has already formulated a kind of test: If any such motivations are indeed the source of moral principles then they should make the will of the person acting unqualifiedly good. The principle of the shopkeeper is to act in his self-interest, so if moral principles are indeed those that serve self-interest, those actions should have genuine moral worth. In each case, however, Kant aims to point out that they obviously do not. It is only when the action is based on a choice whose principle gives that choice unqualified value that the action expresses an unqualifiedly good will and so has moral worth. The motive "It is my duty" is just a stand-in for this principle, and its value is quite different

from self-interest and the rest of those motivations. Or so I think we should understand Kant's discussion.

Indeed, a pretty clear factor in favor of my understanding is Kant's own statement about these examples: "It is clear from our previous discussion that the objectives we may have in acting, and also our actions' effects considered as ends and as what motivates our volition, can give to actions no unconditional or moral worth" (G 4:400). He seems to be saying just what I have proposed, that in none of the examples he has given us is the will unqualifiedly good until the example is changed so that the person's choice is based on duty. In each case, the value of the choice is qualified by the value of some outcome or intended outcome. The shopkeeper's decision is good for his self-interest, but achieving one's own interest itself isn't always or unqualifiedly good. Even in the case of the person who helps out of sympathy, the value of her action depends on the value of her choice based on sympathy. Sympathetic choices are of value because a sympathetic maxim aims at furthering the welfare of some other person. But aiming or achieving that is not always worthwhile. As Barbara Herman points out,

Suppose I see someone struggling, late at night, with a heavy burden at the back door of the Museum of Fine Arts. Because of my sympathetic temper I feel the immediate inclination to help him out. . . . We need not pursue the example to see its point.¹⁶

I might put the point in a slightly different way: The value of the welfare of the person at which her sympathy aims is a condition of the value of her choice to help out. But in some circumstances, that person's welfare is not good, for instance, a circumstance in which promoting it requires violating another person's rights. In that circumstance, a condition under which the choice to help would be good (that the promotion of the recipient's welfare is of value) fails to obtain. Indeed, any choice based on a sympathetic principle can be imagined in different circumstances in which the condition under which it is good does not obtain. So a choice based on that principle is not unconditionally good. The same is true in the other examples as well: choosing to preserve one's life based on the fear of death, and choosing on the basis of a natural inclination to happiness to prudently secure the means to it in the future. The value of the choices each person makes is conditional on the value of what the choice aims to accomplish. Since whatever motivates a good will is that principle which will serve as the fundamental principle of morality – the principle distinguishing right from wrong – none of the bases of these alternative choices can be that principle, since in none of these cases is a choice based on those principles good without qualification.

Now a key feature of this reading of Kant's examples is that it supposes that actions in some sense do or at least can exemplify the principles motivating them.¹⁷ That is what draws the response of respect in us for the will of the agent and thereby grounds our judgment of the moral worth of the action. It is the expression of the moral law in the action of the person acting from duty that draws our esteem. It will be helpful to try to be explicit about what "expressing a principle"

might amount to. Let's suppose that an action expresses the principle that motivates it by, in some sense, *exemplifying* those principles. Thus, "acting on principle" expresses the principle on which you act in such a way that its expression engenders the appropriate attitude for the principle, so expressed, in others.

First off, an action that expresses a principle that motivates it must at least *conform* to the principle that motivates it. This would explain why Kant begins his discussion of moral worth by "passing over" "all actions that are already recognized as contrary to duty" (G 4:397). What he says is that "the question whether they might have been done *from duty* never arises." But we can also see that such actions could not in any case *exemplify* the moral principle on which they might be based (and so could not be objects of the relevant respect reserved for such principles). When someone does the wrong thing for the right reason, or the right thing for the wrong reason, as we say, their actions belie the principles that motivate them. Their actions do not exemplify their motivating principles. Let us assume that an action conforms to a principle when the act-description in that principle accurately describes the agent's action, and an action will be accurately described by a principle when it possesses the relevant features referred to in the principle. For instance, if a principle says to tell the truth and an act possesses the feature of being a truth telling then the action conforms to the principle.

Now many of our acts of speaking are truth tellings, just as much of our behavior trivially conforms to many other principles. However, little of that behavior actually *expresses* a principle. This is because that behavior does not provide an *example* of the principle to which it conforms. Imagine how one might use a bit of behavior as an example to explain a principle to someone, such as a child or someone learning a skill. One would not explain the principle of telling the truth by pointing to someone reading aloud out of an encyclopedia, for instance, or simply talking accurately about this or that set of facts. Nor would a scoutmaster point to just any person helping another to cross the street as an example of the scouting rule "be kind." In each case, it would be insufficient to provide an example of a principle to pick out someone who was merely conforming to the given principle or rule.

We can understand more precisely what is required for exemplification of a principle by drawing on some elements from Nelson Goodman's account of artistic expression, which involves, as he puts it, a kind of symbolic back-reference to the thing exemplified.¹⁸ Goodman employs the example of a tailor's color swatch.¹⁹ Not just anything of a particular hue of blue exemplifies that hue; a tailor's swatch, in addition to *being* blue, itself *refers* back to that blue. Likewise, my suggestion is that not just any action conforming to a principle exemplifies the principle. The action must additionally refer back to the principle to which it conforms, much as the tailor's color swatch not only *is* blue, but *refers* to that blue. Thus, an element beyond conformity to principle is required, something to make the action refer back to the principle or rule to which it conforms.

Color swatches refer back to the colors they possess because they are elements of a symbolic system within the tailoring practice in which they play this role. Perhaps they acquire this referential role because there exists a practice in which

tailors use them to refer to a fabric color. Although actions that are performed on principle exemplify the principle, moral action is not in this way a part of a system of symbols (as it would be if, for instance, it were part of a ritual or theatre). But action on principle does refer back to the principle to which the action conforms. In teaching a principle, explaining what the principle is, and so on, we appeal to such actions, and not to actions that merely conform to the principle. The actions that exemplify a principle conform to a principle *because* the agent has chosen on the basis of that principle. Thus willing an action that conforms to a principle on the basis of the very principle to which it conforms refers back to the principle. The agent intends that his action conform to the principle, and in so intending, make his action an example of the principle. That reference back to the principle is missing when an action was not motivated and justified by the principle to which it conforms. Since actions cannot literally be examples of principles, however, I take it that they are metaphorical exemplifications or, as Goodman would term it, *expressions* of principles.

8. Overdetermination and Expressing the Moral Law

Obviously more needs to defend the sense of “expression” I delineated in the last section, but I hope it is clear enough so that from hereon we can simply assume that something such like this is the case in actions that are done *because* they conform to a practical principle. They express the principles on which they are based by, metaphorically speaking, exemplifying those principles. I take it the “because” in “acting because of conformity with a principle” includes both an explanatory and justificatory sense. But we can imagine that what explains a person’s action (that is, motivates it) is one thing, while how she justifies it can be quite another. At any rate, that is how things often appear, so I shall assume that this does on occasion happen.

Now, suppose in some case someone’s action is motivated by one principle that does not justify her action, but she regards it (correctly, we will assume) as justified by quite another principle. What motivates her to tell the truth in some case is the principle of telling the truth when there is a chance of being caught (adopted, say, because she is highly averse to being caught), but she believes her action is justified because she morally ought to do so. She embraces the moral principle of not lying, let’s say, but that she embraces this principle is not why she tells the truth. She does it out of fear of being caught. Such an action, intuitively, does not fully express her will. It expresses her interest in avoiding embarrassment perhaps, but it does not express her embrace of the principle of not lying – even though it conforms to that principle. Such an action, it seems to me, does not express a good will in the sense in which Kant was talking of it.

Perhaps the reverse situation is conceivable, a situation in which a person justifies her action in non-moral terms but is wholly motivated by moral principle, although it is puzzling to me exactly how such a case might work, and so how to classify such an action. So, as a first approximation, then, let us say that some

person S's action A expresses some (moral or non-moral) principle, M, only when

- (i) A conforms to M,
- (ii) S's acceptance of M explains her choice of A,
- (iii) S's justification for choosing A is that A conforms to M.

For instance, suppose S helps his grandmother around the house. S's helping exemplifies the Boy Scout principle "be kind" only if (A) the help was an act of kindness, (B) S chose to help because S embraced that scouting principle, and, finally, (C) that principle is what, by S's own lights, counted in favor of choosing to help. In a case of this kind, we shall say, S's act, because it expresses this scouting principle, has genuine "scouting worth." Those who likewise embrace scouting principles will, at least if they are not suffering from depression or some other abnormal psychological state, respond with scouting esteem to S's act. The principle they hold in high standing is there before them, exemplified by the action metaphorically speaking, and because that scouting principle is there before them, it engenders a response of esteem in them.

For moral principles, the same will apply *mutatis mutandis*. Suppose S helps his grandmother around the house. Suppose "help others" is a moral principle. S's helping exemplifies the moral principle "help others", only if (A) the help was an act of helping, (B) S chose to help because S embraced that moral principle, and, finally, (C) that principle is what, by S's own lights, counted in favor of helping. In a case of this kind, we shall say, S's act, because it expresses this scouting principle, has genuine "moral worth." Those who embrace that moral principle will, at least if they are not suffering from depression or some other abnormal psychological state, esteem S's act. The principle they hold in high standing is there before them, exemplified by the action, metaphorically speaking, and because that scouting principle is there before them, it engenders a response of esteem in them.

If my portrayal of this so far is correct, then moral worth is just a kind of value that is explained in the same way as many other sorts of value. For any system of practical rules (clubs, governments, orders) there will be actions that exemplify those rules, and those actions will have the relevant "worth" ("scouting worth", "civil worth", "military worth" etc.). The difference is that *moral* worth is a special kind of value because of the special nature of the principle that is expressed in action that is chosen on its basis, because of the supreme importance of the moral law and the special esteem we have for that law.

This leaves at least two important questions that need answering in order to be more precise about what acting from duty is. First, in order to express M, must S's acceptance of M *wholly* explain and be S's justification for A, or can these things cooperate with other principles motivating or favoring A? Second, practical principles of any kind can presumably require, forbid or permit a given action. Can an action that is not required by a practical principle, but is, say, only recommended, have any worth? The scouting rule "Be kind" for instance, requires us to help people struggling to cross the street, forbids us from making fun of someone's

shortcomings, but may permit us to stay home and watch TV rather than work at the soup kitchen any given evening. Suppose M merely permits rather than requires A. If M does not *require* that S do A, then how can S's acceptance of M explain and justify her doing A? For instance, suppose we decide to go help at the soup kitchen this evening, though going is not required by the principle. In what sense might it be because we embrace the scouting rule "Be kind" that we do so? Does such an action have any scouting worth?

Consider these schemas of examples to answer these questions. Assume agent S (e.g., a boy scout) does some action A (e.g., tells the truth) that conforms to some principle M ("Be truthful").

Example 1:

- (i) M requires that S do A,
- (ii) S's acceptance of M motivates her choosing to A,
- (iii) S's justification for choosing A is that M requires it.
- (iv) No other principle to choose A motivates S except M.

For instance, suppose the scouting principle "Be truthful" requires a boy scout to tell the truth to his parents on some occasion. And suppose it is the scout's acceptance of that principle as a guide to behavior that motivates him to tell the truth. Moreover, the principle that justifies his telling the truth, in his eyes, is that the scouting principle requires it of him. And finally, although the scout may recognize that other principles would counsel him to tell the truth, they play no role in his reasoning to do so. It was simply that scouting requires it.

Example 2:

- (i) M requires S to A.
- (ii) S accepts some principle, N, to A other than that M requires it.
- (iii) S's acceptance of M and N are together necessary to motivate her to A.
- (iv) S's justification for A is that both M requires it and that N requires it.

Continuing with our scouting analogy, suppose the scout also knows that his parents are very good at detecting his lies, and offer stern punishment for lying. But S only cares some about avoiding punishment, and only cares some about the scouting rule "Be truthful". Alone, neither is enough to get him to tell the truth. But together, they weigh in favor of doing so, and so he tells the truth. Moreover, in justifying his actions to others, he will say genuinely that he tells the truth because together M and N are sufficient to make sense of doing so.

Example 3:

- (i) M requires that S do A.
- (ii) S accepts some principle, N, to A other than that M requires it.

- (iii) S's acceptance of M alone motivates her choice to A.
- (iv) S's lone justification for choosing to A is that M requires it.

In this case, while he knows of the threat of punishment, it plays no role in motivating or justifying his action. It is only his embrace of the scouting principle that does. However, had M not required S to A, he would have done so anyway, because N also required it.

Example 4:

- (i) M requires that S do A.
- (ii) S accepts some principle to A, N, other than that M requires it that is sufficient to both motivate and to justify S's doing A.
- (iii) S's acceptance of M and N motivates her choice of A.
- (iv) S's justification for choosing A is that M and N require it.

In the final example, we have a classic case of "over-determination" of an action by motives. Unlike the second case, in which both principles are jointly sufficient for S to arrive at the choice, in this case, both principles are independently sufficient to arrive at it, both in the sense that they motivate and justify the action. Our scout, then, tells the truth both because he embraces the scouting principle *and* he fears his parents' retribution.

There are other schemas to consider, but these are enough to further the discussion. First note that in example 4, two contrary to fact conditionals are true: If the scout had not embraced the scouting principle "Be truthful", then he still would have told the truth (out of fear of punishment). And, by the same token, if the scout had not feared punishment, he still would have told the truth (because of his embrace of the principle). This is a difference between example 4 and example 2. In example 2, had he not been somewhat afraid of punishment, he would not have told the truth. Likewise, had he not embraced to some extent the principle "Be truthful", he would have lied. In case 4, unlike in case 2, he did not need the threat of punishment to conform to the principle. Likewise, in cases 1 and 3. In case 3, if there were no threat of punishment, it would make no difference to whether he was truthful, and in case 1, there is no such threat.

Now the actions that possess what we shall call "M worth" – that is, the sort of value I described as based on the fact that a given action expresses some practical principle or rule – are those which express the volitional principle M characterizing that will. Actions express their motivating principles when they exemplify those principles, metaphorically speaking. And they do this, as I have described above, when those actions not only conform to those principles, but also contain a reference back to the principle to which they conform. This is the way in which a swatch of fabric exemplifies a color, for instance. In the case of actions, the reference to the principle is contained in the fact that the person is conforming to the principle *for the sake of the principle* itself.

It seems to me that 1 and 3 are certainly the sorts of actions that would be expressions of principle M. They are the sorts of actions one might, for instance, use as examples, to explain or teach the principle, for instance. In example 2, by contrast, M is not expressed by the action, even if the action in 2 is explained in part by the fact that the agent embraces M. It would not be a genuine expression of the scouting principle, “Be truthful”, for instance. It seems, instead, to express a will that half-heartedly embraces both a fear of punishment and scouting principles, such as “Avoid telling falsehoods” and “Avoid punishments.” It is not the sort of action one would explain or teach the principle M with, and would not engender the esteem of those who hold this scouting principle dear. Finally, although in 4 the agent’s embrace of M is sufficient for his action, in the sense that even if there were no other cooperating principles, he would have conformed to it, it is not clear whether the action expresses M or the cooperating principle because we don’t stipulate which was the “real” explanation and justification for his conforming to M.

Suppose, then, that moral worth is simply moral goodness of the good will expressed in actions. And suppose further that moral goodness is being good in any circumstance. Then when M is the moral law, 2 and 4 do not express something that is valuable in every circumstance. In example 2, it is because the principle expressed would not be the moral law alone; it would be the moral law together with, for instance, fear of punishment. In example 4, however, there is no unqualified goodness expressed, not because what is expressed has conditional value, but because it is not an expression of any single principle of volition; it is an expression of two principles, and so is opaque. But it doesn’t follow that 2 and 4 are bad actions. They are good in some, but not in all, respects.

As I see it, then, only 1 and 3 are examples of what Kant thinks of as “acting solely from duty”. They turn out to be instantiations of a general schema that applies to a wide range of practical principles and actions (that can exemplify the principles on which they are based). Add this to the thesis Kant holds about our reaction to the moral law, when it is exemplified in a person’s behavior, and we arrive at the following conclusion: Actions have moral worth when the principle they exemplify is the moral law, and this is so because the judgment of “moral worth” is nothing over and above our own reaction of respect toward that law, so exemplified.

9. Conclusion: From the Good Will to the Fundamental Principle of Morality

The connection between the good will, the moral worth of actions from duty, and the first formulation of the Categorical Imperative emerges once we properly understand what Kant is up to in each discussion. To be valuable is to be the object of a rational will. To be unqualifiedly valuable is to be unqualifiedly the object of a rational will. And to be an unqualifiedly valuable object of a rational will is to be willed by rational agents, insofar as they are rational, in any

conceivable circumstance. The principle that is so willed will be expressed in the actions of a person with a good will. The good will in such actions expresses what is the object of rational willing under any conceivable circumstance, the moral law, and hence express what is unqualifiedly good.

The search in the *Groundwork* is for a motivating principle such that, when acted on, it would “make my volition morally good”, that is, good without qualification. That, Kant assumes, must be the fundamental principle of morality, the principle he is trying to discover and justify. For it follows from the fact that a given choice has unqualified value that it is good because of some intrinsic property it has. But the intrinsic property of any choice is just the principle that motivates and justifies that choice. It is an extrinsic property of a choice that it furthers self-interest, sympathetic and eudemonic ends. Choices based on those principles are good because they further our interests, promote others’ welfare or our own overall happiness. In each case, we can imagine circumstances in which the choice based on such principles is not good because what is furthered by such a choice is not good. That raises this question: If we want to find out what makes a will unqualifiedly good, since we must focus on the intrinsic property of the choice, its principle itself, and not what comes of that choice, such as the consequences of choosing on some principle, what sort of principle could be the object of rational willing no matter what the circumstances? That is, What sort of principle could be that in virtue of which a choice is unqualifiedly good? But then we are focusing simply on the form of that principle, not what it offers as an outcome of choosing on its basis.

If what is unqualifiedly good is what I could choose in any conceivable circumstance then the critical question for Kant’s project is, Which principle could I rationally choose that I act on in any conceivable circumstance? I end with just Kant’s own words, for Kant takes this question to be the same as asking

“Can you will that your maxim become a universal law?” If not, that maxim must be repudiated . . . because it cannot fit as a principle into a possible universal legislation, and reason forces me to offer my immediate respect to such legislation (G 4:403).

Notes

- 1 All except one, the goal of moral perfection, a goal the achievement of which would result in, naturally, having a good will.
- 2 Kant often refers to the human will with one umbrella term, *Wille*. But in fact he distinguishes two capacities implicit in the human will, *Willkür* or the capacity of choice (the capacity shared with animals), and (using the term in a narrower sense than above) *Wille*, the practical capacity of reason itself, the capacity to lay down laws of behavior for us.
- 3 There is quite a lot of difference of opinion about what exactly a maxim is, and, indeed, whether we are acting on a practical law when we act on a maxim that could be a principle every rational agent could act on. See, e.g., Allison, 1990.
- 4 I am here reading Kant as thus to some degree in agreement with T. M. Scanlon and others who are “buck-passers” about the good. See T. Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998), 95–100.

- 5 *Principia Ethica* (1903; rev. ed., Cambridge: Cambridge University Press, 1993), 78–80.
- 6 Michael Zimmerman argues this in Zimmerman 2001, p. 145.
- 7 Some inventors apparently work this way. In their mind's eye, they construct a given device, and then imagine how it would operate. When they first do it, they discover, in imagination, that there is a hitch, some gear missing, some different fiber that needs to be used, and so on.
- 8 For a notable exception, see Foot 1997, pp. 172–4.
- 9 Henson 1979, pp. 39–54. See also Herman, 1981, pp. 359–82.
- 10 See Johnson 1996, pp. 313–37, for a fuller discussion of some of these ideas.
- 11 For this reading, see, e.g., Wood, 1999, pp. 27, 30–33
- 12 “Before a humble common man in whom I perceive uprightness of character in a higher degree than I am aware of in myself *my spirit bows*, whether I want it or whether I do not and hold my head ever so high, that he may not overlook my superior position. . . . Respect (*Achtung*) is a *tribute* that we cannot refuse to pay to merit, whether we want to or not; we may indeed withhold it outwardly but we still cannot help feeling it inwardly” (CPrR 5:77).
- 13 “What I recognize directly as a law for myself, I recognize with respect, which means nothing more than the consciousness of my will’s *submission* to the law, without the mediation of any other influences on my mind.”
- 14 Both Henson and Herman think this idea is key to understanding Kant’s position.
- 15 Herman, 1981. See also Baron, 1984, pp. 197–220.
- 16 Herman, 1981, pp. 364–5; see also Baron, 1984.
- 17 “All reverence for a person is properly only reverence for the law (of honesty and so on) of which that person gives us an example. Because we regard the development of our talents as a duty, we see too in a man of talent a sort of *example of the law* (the law of becoming like him by practice), and this is what constitutes our reverence for him” (G 4:401n).
- “ . . . for their worth consists . . . in the attitudes of mind – that is, in the maxims of the will – which are ready in this way to manifest themselves in action . . . they exhibit the will which performs them as an object of immediate reverence” (G 4:435).
- 18 Thanks to Nelson Potter for suggesting that I consider Goodman’s account.
- 19 Goodman, 1976, Ch. 2.

Bibliography

- Allison, H. 1990: *Kant’s Theory of Freedom*. Cambridge: Cambridge University Press.
- Baron, Marcia. 1984: The alleged repugnance of acting from duty. *Journal of Philosophy* 81: 197–220.
- Foot, P. 1997: Virtues and vices. Reprinted in R. Crisp and M. Slote (eds.), *Virtue Ethics*. Oxford: Oxford University Press.
- Goodman, N. 1976: *Languages of Art: An Approach to a Theory of Symbols*, 2nd edition. Indianapolis: Hackett Publishing Company.
- Henson, R. 1979: What Kant might have said: Moral worth and the overdetermination of dutiful action. *Philosophical Review*, 39–54.
- Herman, B. 1981: On the value of acting from the motive of duty. *Philosophical Review*, 359–82.
- Johnson, R. 1996: Kant’s conception of merit. *Pacific Philosophical Quarterly* 77.
- Wood, A. 1999: *Kant’s Ethical Thought*. Cambridge: Cambridge University Press.
- Zimmerman, M. 2001: *The Nature of Intrinsic Value*. Lanham, Md.: Rowman & Littlefield.