

A KANTIAN ACCOUNT OF HUMAN VIRTUE

by

Christopher Anthony Brown

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PHILOSOPHY

In Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2008

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Christopher Anthony Brown entitled A Kantian Account of Human Virtue and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

_____ Date: 4/16/2008
David Schmitz

_____ Date: 4/16/2008
Houston Smit

_____ Date: 4/16/2008
Mark Timmons

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: 4/16/2008
Dissertation Director: David Schmitz

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Christopher Anthony Brown

TABLE OF CONTENTS

ABSTRACT.....	5
INTRODUCTION.....	7
1. VIRTUE AS APTITUDE FOR LAWFUL CONDUCT.....	20
2. NATURAL MORAL GOODNESS.....	44
3. VIRTUE AND PRACTICAL LAWS.....	59
4. ONE'S OWN MORAL PERFECTION.....	70
5. ONE'S OWN HAPPINESS.....	79
6. THE MORAL PERFECTION OF OTHERS.....	87
7. VIRTUE AND HONESTY.....	113
8. HUMAN VIRTUE AND CONCERN FOR NON-AGENTS.....	126
REFERENCES.....	159

ABSTRACT

There are certain elements of Kant's moral philosophy that I believe no moral theory can afford to ignore. On the other hand, there are others which Kant's theory evidently would be better off without. I will be developing an account of human virtue by defending and exploiting some of Kant's most fertile and sustainable ideas, while arguing against other theses of his, a few of which have come to be regarded as definitive of Kantian Ethics.

I begin by showing that we can plausibly interpret Kant's texts on “the good will” and “actions from duty” as presupposing that an agent's moral goodness consists in her aptitude for lawful conduct, that is, her aptitude for living in accord with practical principles valid for all possible agents. I build my basic account of virtue by showing that this aptitude inheres in the possession of certain traits. The cornerstone of virtue, I argue, is the moral commitment: the stable, non-instrumental aim of living lawfully. For, when this commitment prevails in determining an agent's actions, lawful conduct necessarily ensues—which cannot be said of any other commitment or aim. I identify four additional elements of virtue by searching for traits that a perfectly morally committed agent might possess, and which together would guarantee that her moral commitment prevails. These are: moral understanding, strength of will, empirical understanding, and empirical power. And I suggest that when an agent violates a moral requirement, this is always saliently attributable to a lack of one or more of the five proposed elements of virtue.

I supplement the basic account of virtue by arguing that the morally committed human agent is rationally required to adopt four further, general aims: the efficient pursuit of her own happiness, the happiness of other agents and non-agents, her own moral perfection, and that of other similarly committed agents. This part of my view differs significantly from Kant's, so I will be largely concerned with critiquing the relevant arguments of his. But the result is still very much a Kantian account, and one that warrants serious consideration in contemporary debates about virtue.

INTRODUCTION

In 2004, David Copp and David Sobel wrote, “These are boom years for the study of the virtues.”¹ They were surely right. The last few decades have seen a lot of growth in the study of morally good character. Arguably, the traditional Greek accounts have not enjoyed this much popularity and scholarship in centuries. And there is no shortage of novel accounts, either. Among these are even those which combine *consequentialist* foundations with the emphasis on character. But comparatively little attention has been given lately to squaring distinctively Kantian thought with questions about what makes a human agent a morally good agent. When Kant *is* mentioned in discussions of virtue, this is usually either to contrast his moral theory with “Virtue Ethics” or to clarify his declared conception of virtue. As several Kant scholars have pointed out, the contrast is not as sharp as it might seem.² But even though Virtue Ethicists are probably becoming more and more aware of this, I know of none who has endorsed a Kantian account of virtue as a serious contender. This is to no small extent understandable. After all, Kant endorses a number of very controversial theses in his moral philosophy, e.g. that lying is never morally permissible; that non-rational beings are *mere means*; and that pursuing one’s own happiness is morally worthwhile only when one’s ability to fulfill one’s duties

¹ Copp and Sobel, “Morality and Virtue: An Assessment of Some Recent Work in Virtue Ethics,” *Ethics*, 114 (April 2004), 514-554.

² See, for example, Robert Louden’s “Kant’s Virtue Ethics,” *Philosophy*, 61 (Oct., 1986), 473-89, or Nancy Sherman’s *Making a Necessity of Virtue: Aristotle and Kant on Virtue* (Cambridge: Cambridge University Press, 1997).

is at risk. So, those who are after a plausible account of human virtue might reasonably suspect that any *truly* Kantian approach would unavoidably come with more philosophical baggage than it's worth. Part of what I want to do in the next eight chapters is put such suspicions to rest, by arguing (among other things) that Kant really is not entitled to the above theses. But I will do this by exploiting certain foundational elements of Kant's view—ideas which, I argue, should play a central role in *any* account of virtue. And I contend that the account developed here is not only an improvement on Kant's own, but one that deserves serious consideration in debates about virtue, even from those who count themselves as opponents of Kantianism.

In the first three chapters, I develop and defend a basic account of virtue as *aptitude for lawful conduct*. Roughly, this view has it that an agent's moral goodness consists in her possession of traits which enable her to live in accord with practical principles that are valid for all possible agents.

I came to this view while trying to figure out why Kant might be right to think so highly of *the good will* and actions *from duty*. In the *Groundwork of the Metaphysics of Morals*, Kant claims that only the good will “is good without limitation” (*Gl*, 4:393), and that only actions from duty admit of “moral worth” (*Gl*, 4:400-1).³ Unfortunately, he

³ Immanuel Kant, *Groundwork of the Metaphysics of Morals*, in Mary Gregor (ed. and trans.), *Practical Philosophy* (Cambridge: Cambridge University Press, 1996). Unless otherwise noted, I will use Gregor's text for translations of Kant's *Groundwork of the Metaphysics of Morals* (abbreviated as *Gl*); *Critique of Practical Reason (KpV)*; and *The Metaphysics of Morals (MdS)*. For translations of Kant's *Critique of Judgment (KU)*, I use James Meredith (trans.), *Critique of Judgment* (New York: Oxford University Press, 1952). And for Kant's *Religion within the Boundaries of Mere Reason (Rel)*, I use Wood & Giovanni (eds. And trans.), *Religion within the Boundaries of Mere Reason* (Cambridge: Cambridge University Press, 1998). Page references for all Kantian texts will use the volume and page number(s) from *Kant's gesammelte Schriften*,

never explicitly gives his reasons for thinking this. I eventually came to suspect that the answer was connected to something Kant says in the Preface of the *Groundwork*:

For, in the case of what is to be morally good it is not enough that it *conform* with the moral law but it must also be done *for the sake of the moral law*; without this, that conformity is only very contingent and precarious, since a ground that is not moral will indeed now and then produce actions in conformity with the law, but it will also produce actions contrary to the law. (*Gl*, 4:390)

Kant's message here, on my reading, is roughly that when an agent's primary aim is *something other than* conformity with the moral law, then it is much more a matter of coincidence whether she acts/lives in conformity with the moral law. Kant may have underestimated how often other aims lead to lawful conduct. But at the very least we can say that when an agent aims to act/live lawfully, and achieves what she aims at, she *necessarily* acts lawfully. In other words, lawful conduct is a necessary result when the aim of acting/living lawfully prevails. And, as I argue in Chapter One, this cannot be said of any other aim. And because this aim is present in the good will and in actions from duty, as Kant defines them, this might explain why he thought so highly of them.

In any case, this is one of my reasons for thinking of the stable, fundamental aim of living lawfully—*the moral commitment*, as I call it—as the cornerstone of virtue. When this aim prevails, lawful conduct is the necessary result. And if we want to maintain that the (ideally) virtuous agent *never fails* to act lawfully, and that this is not due to coincidence, then we *must* count the moral commitment as a necessary constituent of virtue.

edited by the Royal Prussian Academy of Sciences (Berlin: Georg Reimer, later Walter de Gruyter & Co., 1990-).

But the virtuous agent must also be endowed with traits that ensure that her moral commitment *does* prevail. I argue for four additional elements of virtue:

- Moral understanding: Understanding the universally valid practical principles and what they imply about how one should live (given one's contingent circumstances).
- Strength of will: The capacity to resist episodic temptations to neglect the ends one values most.
- Empirical power: The capacity to bring about the things one intends to bring about.
- Empirical understanding: Understanding one's world, including one's own psychology.

I argue that the agent who has each of the five elements in its fullest form never fails to act lawfully, and that our failures are always due to a deficiency in one or more of them. Thus, I propose, an agent's moral goodness—again, her aptitude for lawful conduct—is determined entirely by the extent to which she has each of the five elements.

In Chapter Two, I defend two key differences between Kant's account of moral goodness and mine. In *Religion within the Boundaries of Reason Alone*, Kant argues that an agent is not morally good unless (a) her choice to aim at living in accord with moral requirements is independent of natural causes, and (b) this aim is her *sole* stable and fundamental aim (*Rel*, 6:21-5). But Kant's argument for (a) presupposes that an agent's character must be *imputable to her* (rather than to nature), if she is to be regarded as

morally good. And his argument for (b) presupposes that, if one has a fundamental (or, non-instrumental) aim of living in accord with moral requirements, one cannot possibly have further fundamental aims. I argue that both presuppositions can plausibly be rejected, and that the resulting account allows us to make more fine-grained moral appraisals of agents. On my view, an agent admits of at least some degree of moral goodness, or, virtue, if she has the moral commitment—no matter how strong her commitment is and no matter how she has acquired it. This is because the moral commitment endows her with an aptitude for lawful conduct that is absent in agents who altogether lack this commitment, all other things being equal. (Their lawful conduct is more a product of coincidence than hers.) In contrast, Kant's view seems to force us to appraise all agents who have not adopted the *perfect* moral commitment as *equally lacking in moral goodness*.

All of this still leaves unanswered the question of why an account of virtue should emphasize *lawful conduct*, as I (following Kant) have defined it. One could reasonably ask, Why think that the virtuous *human* agent must be committed to living in accord with practical principles valid for all *possible* agents? The moral commitment could have been defined otherwise. Why is it not enough, for example, if a human agent concerns herself with identifying and living in accord with principles valid for all *human* agents? I have two answers, which I present in Chapter Three. I begin by defending and then employing the Virtue-Ethical canon that the virtuous agent does the right things for the right reasons. I argue that when we apply this canon to the virtuous agent's moral

thinking, we get the result that she grounds her beliefs about how she should live on her beliefs about principles' validity for all possible agents. In this sense she thinks like a Kantian philosopher and, I contend, so should any human agent who sincerely aims to approximate virtue. To be sure, though, a rather different kind of argument is needed to show that the principles valid for all possible agents would have the robust *motivational* role I claim they have. Why should accord with these principles be the focus of the morally good human agent as she is figuring out which (overt) actions to perform? I will show that if we define the moral commitment in a more human-centered way, then even the *ideally* virtuous agent's accord with moral requirements is to a great extent dependent on things that are external to her character, namely, the features of the world she calls home. My account, on the other hand, enables us to maintain that the ideally virtuous agent *always* does what is morally best—whether she lives in our world or any other.

It may seem that this theoretical perk comes at too high of a cost. Once we properly understand the universal nature of the moral commitment, the thesis that the moral commitment is *necessary* for virtue seems to have the unattractive implication that there are very few human agents who admit of any degree of virtue. But I suggest that there are probably many human agents who are morally committed, according to my definition, but who do not claim, or even recognize, themselves as such. We can see this if we just imagine people whom we sincerely regard as morally good being thrust into radically different, *other-worldly* situations, and ask ourselves how we'd expect them to respond.

This will conclude my defense of the basic account of virtue as aptitude for lawful conduct. Although this much of the view to be developed here is not shy about its Kantian roots, my hope is that Kantian sympathies are not needed to appreciate the arguments in the first three chapters. Of course, I also hope that I will not have lost too many card-carrying Kantians along the way. The basic account is intended to be neutral, at least in the sense that it says nothing about the *content* of the practical principles valid for all agents—*practical laws*, as I will call them. For all I will have said by the end of Chapter Three, there may just be one practical law, and it could be Eudaimonistic (“Flourish!”), consequentialist (“Maximize the aggregate of happiness in your world”), Kantian (“Always treat rational nature as an end, and never merely as a means”), or who knows what else. I suspect, though, that there are more than one. My moral understanding being as far from perfect as it is, I do not have a complete account of them. However, at several points between the beginning of Chapter Four and the end of Chapter Eight, I will argue against some of Kant’s claims about such laws. For example, Kant claims that there is a practical law which prohibits lying, and that his Formula of Universal Law and his Formula of Humanity are one and the same practical law. I address these claims in chapters Seven and Eight, respectively.

But my primary objective after Chapter Three is to further fill out the basic account, as it applies to *human* agents. I argue for four general aims that, I claim, morally committed human agents should adopt and maintain. Two of these are *self-regarding*. In Chapter Four, I side with Kant in claiming a duty to aim at promoting one’s own moral

perfection. To my knowledge, Kant never provides an argument for this claim. But I argue that it would be irrational for a morally committed human agent *not to* aim at morally improving herself. As a morally committed agent, she aims to live lawfully. But as a *human* agent, she should realize that she can always improve her aptitude for lawful conduct, at least by increasing her moral understanding, empirical understanding, and empirical power. (I am not quite sure that a human agent can deliberately improve her moral commitment or her strength of will.)

In Chapter Five, I argue for a similar aim concerning one's own *happiness*. Kant disagrees. He does not say that there is no duty at all to promote one's own happiness, but he calls this an *indirect* duty to signify that the morally legitimate end served by promoting one's own happiness is something other than one's happiness itself. Roughly, Kant holds that being unhappy leads to transgressions of duty, so we are to avoid slipping into this state. I, on the other hand, will be arguing for a *direct* duty. As we will see, Kant offers two arguments against such a duty. One has it that duties must be concerned with aims that we are reluctant to adopt; and since all human agents have their own happiness as a natural end, it is contradictory to suppose that there is a duty to promote one's own happiness (*MdS*, 6:385-6). He also argues that there can be no practical law of self-love, that is, one requiring agents to promote their own happiness. This would have to be a "determinate" practical principle which is valid for all possible agents, says Kant. And no such principle is possible, since, for one thing, agents differ in what they need in order to be happy; for another, an agent would have to be omniscient in order to know

what *she* needs to be happy (*Gl*, 4:418-9; *KpV*, 5:19-26). But I will argue that Kant cannot consistently require practical laws to be determinate in the above senses, since he claims at least one practical law which is not. Moreover, there seems to be no good reason to impose this requirement on practical laws in the first place. I concede that I am forced to remain agnostic on the existence of a practical law of self-love. But I show that a duty concerned with the *efficient pursuit* of one's own happiness can be derived from a very different, plausible practical law. Moreover, I show that this duty satisfies Kant's reluctance requirement on duties.

Having argued that the morally committed human agent should have the above two *self-regarding* aims, I work toward similar *other-regarding* aims in Chapter Six. In the *Groundwork*, Kant uses the Formula of Universal Law (FUL) to argue for a duty to promote other agents' happiness. FUL states that we are to act only on those maxims that we can will as universal laws. And Kant argues that no human agent can will a maxim *not to* promote other agents' happiness as a universal law. Since it is, according to FUL, impermissible to act (omit) on such a maxim, it follows that we should instead be willing to promote others' happiness (*Gl*, 4:423). I not only think FUL is a practical law, but also that Kant's use of it in the above argument is correct. So, Chapter Six has very little to add on this score. (Chapter Eight, on the other hand, argues that Kant is wrong to think this duty is owed *only* to other *agents*.) However, Kant denies that we have a duty to promote other agents' moral perfection. His argument is roughly that, since I cannot make it the case that another person chooses to live in accord with moral requirements, I

cannot have a duty to aim at this (*MdS*, 6:386). Even granting this, however, Kant's conclusion does not follow. For, as I will show, it is possible to morally improve those who have made the aforementioned choice, by improving their moral understanding. And I present arguments, one of which is modeled on the above FUL argument, for a corresponding duty. Analogous arguments can be given for duties, suitably qualified, to aim at promoting other elements of other morally committed agents' virtue, e.g. their strength of will, empirical knowledge, and empirical power. The result is a general duty concerning the moral perfection of others. However, I point out that the duty thus established is, in an important sense, less general than the duty concerning other agents' happiness. The latter is owed to *all* human agents, whereas the former is owed only to *morally committed* human agents.

So, on my view, there are four general aims that the (more or less) virtuous human being has in addition to the moral commitment itself: her moral perfection; the moral perfection of other morally committed human agents; the efficient pursuit of her happiness; and the happiness of others. (Again, the latter aim will be broadened in Chapter Eight.) The question then is whether there is any necessary constituent of human virtue that cannot be accounted for by combining this taxonomy of aims with the basic account of virtue as aptitude for lawful conduct. Consider a human agent who has the four aims and who is as good as a human being could be as far as the moral commitment, moral understanding, etc., are concerned. What might this *nomologically* virtuous human agent lack? My sense is that any trait that comes to mind will fall into one of two groups.

Traits she would in fact possess *because* she has the above traits; and traits that, though intuitively admirable, really are not defensible as *necessary* constituents of human virtue. Of course, I cannot survey all candidates and the arguments that might be given for them. But Chapter Seven (“Virtue and Honesty”) should give the reader some idea of how I would approach other candidate traits.

There I argue against Kant’s famous claim that lying is *always* morally impermissible. As we will see, each of Kant’s arguments against lying presupposes an extremely exaggerated view concerning the frailty of the general believability of statements; Kant overestimates the adverse effects both of lies and of universalized maxims prescribing them. I concede that, given our dependence on each other for information, the best pursuit of the above four aims most often entails refraining from lying (and from other forms of deception). However, I argue that there are real human cases in which truth-telling is flat out wrong and lying is the *morally good* thing to do. So, I contend, the morally committed human agent who has made adequate progress in morally perfecting herself, particularly with respect to moral and empirical understanding, *will not* have an absolutist’s attitude against lying or other forms of dishonesty. Such an attitude would compromise her aptitude for lawful conduct. Instead, she thinks of honesty merely as a mode of conduct that *tends* to be lawful. I conclude that honesty is not a necessary constituent of human virtue.

Finally, in Chapter Eight, I turn to what I found to be the most formidable challenge of all in this project: showing that Kant is wrong to think that there are no direct duties

concerning the happiness or well-being of *non-agents*, e.g. (non-rational) animals. To be sure, Kant is explicit that we have a duty to refrain from cruelty toward animals. But he regards this as an *indirect* duty. On his view, the only morally legitimate end served by kindness toward animals is *one's own perfection* (*MdS*, 6:443). *Direct* duties can be owed only to rational beings, says Kant, since rational nature alone is an end in itself; non-rational beings, on the other hand, are mere means (*Gl*, 4:428-9). According to my diagnosis, a big part of the problem lies with Kant's insistence that *all* of our duties are derivable from the Formula of Humanity (FH), which requires us to always treat rational nature as an end, and never merely as a means (*Gl*, 4:439). As far as I can tell, there is no sound way of deriving any *direct* duty toward non-rational beings from FH. So, if FH in fact *has* the status Kant awards it, there simply *cannot be* any direct duties to non-agents.

However, I will argue in effect that Kant can and should *demote* FH. I will show that his arguments for its status are importantly incomplete, and that they contradict his claim that the Formula of Universal Law is a practical law. Perhaps more importantly, I will also show that Kant's declared (direct) duty concerning other agents' happiness *cannot* be derived from FH. (And notice, by this time we will already have seen, in Chapter Six, that this duty *can* be derived from the Formula of Universal Law (FUL)—which Kant regards as nothing more than an alternative formulation of FH!) Thus, maintaining FH's status renders Kantianism almost entirely indifferent to the happiness, well-being, and suffering of *all* beings, including all agents. For all of these reasons, I contend that the principle many of Kant's defenders have placed at the core of Kantianism should be

given a much less central role. We *should not* think that all of our duties are derivable from FH. And I will show that a direct duty concerning the well-being, if not the happiness, of non-agents can be derived from FUL. But since I follow Kant in thinking of well-being as (at least) a constituent of happiness, I count my results in Chapter Eight as a broadening of the morally committed human agent's aim concerning others' happiness.

CHAPTER ONE

VIRTUE AS APTITUDE FOR LAWFUL CONDUCT

INTRODUCTION

We think of some people as morally good, others as bad or even evil, and yet others as somewhere in between. However, the criteria on which we base our moral appraisals of ourselves and others are not always clear to us. And certainly few of us have the opportunity to spend much time on the question of which criteria we *should* apply in our moral appraisals. But the fact that these appraisals often influence our conduct provides one good, moral reason to acquire the correct criteria, if there are such things. In addition, having the correct criteria in hand is important for those who aim to morally improve themselves. The criteria must, of course, be concerned with some parts or properties of human beings, but which ones should have our attention? Traditionally, Virtue Ethicists have emphasized such intuitively admirable attributes as courage, compassion, justice, generosity, and honesty as constitutive of morally good character. But all the recent press on accounts of virtue evidences that questions about morally good character are no less controversial than questions about morally good action. Contemporary Virtue Ethicists are stepping away from the traditional virtues, and are

looking for novel ways of explaining why their proposed attributes are indeed morally good.⁴

In this chapter I'll be offering a Kantian alternative to the popular accounts of virtue and the morally good character traits, features, and/or parts. Roughly, this view has it that virtue *just is* aptitude for conduct in accord with the moral requirements; and the morally good character traits, etc., are those which augment that aptitude. I identify five traits, which I call the moral perfections:

- The moral commitment: The stable, non-instrumental aim of living in accord with universally valid practical principles.
- Moral understanding: Understanding the universally valid practical principles and what they imply about how one should live (given one's contingent circumstances).
- Strength of will: The capacity to resist episodic temptations to neglect the ends one values most.
- Empirical power: The capacity to bring about the things one intends to bring about.
- Empirical understanding: Understanding one's world, including one's own psychology.

⁴ Among these are Rosalind Hursthouse, *On Virtue Ethics* (Oxford: Oxford University Press, 1999); Thomas Hurka, *Virtue, Vice, and Value* (Oxford: Oxford University Press, 2000); Julia Driver, *Uneasy Virtue* (Cambridge: Cambridge University Press, 2001); Philippa Foot, *Natural Goodness* (Oxford: Oxford University Press, 2001); Michael Slote, *Morals from Motives* (Oxford: Oxford University Press, 2001); and Christine Swanton, *Virtue Ethics* (Oxford: Oxford University Press, 2003).

Only in later chapters will I make any claims about what the universally valid practical principles are. For all I will say here, “Promote the maximum of happiness” could be the only one, but so could “Treat rational nature as an end in itself” or “Live a Eudaimon life”. This is because, for now, I want to keep the account of morally good character as neutral as possible. So, I should point out that when I say that an agent has the moral commitment, I do not mean that she necessarily has a particular set of principles in mind. Rather, she could express this commitment by saying, “I will live in accord with the universally practical principles, *whatever they are.*” The moral commitment itself is intended to be neutral between moral theories.⁵ It is a morally committed agent’s *moral understanding* that determines which doctrine she follows. And questions about what moral understanding consists in are, I believe, where the real fight is among Virtue Ethicists, but I do not need to weigh in here. For, I will argue that (regardless of how we fill out the above definition of moral understanding), an agent that has each of the five moral perfections to its fullest extent necessarily acts in accord with the universally valid practical principles. Moreover, I contend that any failure to accord with those principles is saliently attributable to a deficiency with respect to one or more of the moral perfections. This is why I think the five moral perfections (as they stand) provide useful and sound criteria for the moral appraisal of agents—regardless of which moral theory one endorses. They also serve to provide goals for those who aim to

⁵ I concede, however, that my definitions of the moral commitment and moral understanding are not compatible with views holding that morality has nothing to do with universal validity. I will ignore these views for now, but in Chapter Three I will argue that the ideally virtuous agent would reject them.

morally improve themselves.

This view owes much to Kant's discussion of the good will and the moral worth of actions. So, I will develop it by discussing the relevant Kantian ideas. In Section I, I argue that possession of a Kantian good will entails possession of the moral commitment. In Section II, I propose that the reason Kant calls the good will "good without limitation" is that it is, compared to other traits one might have in its place, uniquely suited for lawful conduct (conduct in accord with the universally valid practical principles). This, I argue, is precisely because the good will entails the moral commitment. (I reject other Kantian explanations.) In Section III, I argue for the other four moral perfections (moral understanding, strength of will, empirical power, and empirical understanding) by searching for other character traits that, when fully realized, guarantee lawful conduct.

I. GOOD CHARACTER ACCORDING TO KANT

Kant obviously thinks that a good will possesses a certain inherent moral goodness which cannot be diminished. He starts the *Groundwork* with the following:

It is impossible to think of anything at all in the world, or indeed even beyond it, that could be considered good without limitation except a good will. Understanding, wit, judgment and the like, whatever such *talents* of mind may be called, or courage, resolution, and perseverance in one's plans, as qualities of *temperament*, are undoubtedly good and desirable for many purposes, but they can also be extremely evil and harmful if the will which is to make use of these gifts of nature, and whose distinctive constitution is therefore called *character*, is not good. (*Gl*, 4:393.)

Here Kant contrasts the good will with certain other character traits, and asserts that these others can be good or bad, depending on whether the agent that possesses them also has a good will. Kant is pointing out that, although we may be inclined to think of these talents of mind and qualities of temperament as *generally* good, their goodness is in fact contingent on the possession of a good will. The implication seems to be that the goodness of a good will, on the other hand, is not contingent. The phrase, ‘good without limitation’, suggests that Kant also thinks of the good will as having *infinite* goodness. Kant may think this, but I think his emphasis here is not that the good will’s goodness is infinite in magnitude; rather, it is that nothing else that might be true of an agent can diminish the goodness she has in virtue of having a good will. This is confirmed by what comes a page later in the *Groundwork*:

Even if, by a special disfavor of fortune or by the niggardly provision of a stepmotherly nature, this will should wholly lack the capacity to carry out its purpose—if with its greatest efforts it should yet achieve nothing and only the good will were left (not, of course, as a mere wish but as the summoning of all means insofar as they are in our control)—then, like a jewel, it would shine by itself, as something that has its full worth in itself. (*Gl*, 4:394.)

Granted, Kant does use ‘worth’ here, instead of ‘goodness’. But, as we will see, Kant uses ‘worth’, ‘goodness’, and their related terms interchangeably.

But what exactly is a good will, and why does Kant speak so highly of it? The above seems to suggest that having a good will involves summoning all of one’s means for some purpose. And notice, the earlier excerpt also mentions purposes: Although understanding, courage, etc., are good for many purposes, they can be evil if the will that “makes use” of

them is not good. These “gifts of nature” would be means that a good will summons for its purpose, provided it has them. Given a lack of them, a good will may “wholly lack the capacity to carry out its purpose”, but, says Kant, this does not diminish its worth. One way of making sense of these ideas is to say that the goodness, or, worth, of a good will is due to (a) its having certain aims (or, purposes) and (b) its summoning all its means for the sake of pursuing of those aims (or, serving those purposes). This mirrors the way Kant thinks about the worth and goodness of actions *done from* duty: What makes them morally good, or, what endows them with moral worth, is (a) their having a certain motive and (b) that motive’s being the sole incentive that produced the action. In fact, it seems to me that Kant intends his discussion of actions from duty to be a vehicle for getting clear on what a good will is and why it has the unconditional goodness Kant attributes to it. After all, Kant’s discussion of actions from duty comes immediately after the following:

We have, then, to explicate the concept of a will that is to be esteemed in itself and that is good apart from any further purposes, as it already dwells in natural sound understanding and needs not so much to be taught as only to be clarified – this concept that always takes first place in estimating the total worth of our actions and constitutes the condition of all the rest. In order to do so, we shall set before ourselves the concept of **duty**, which contains that of a good will though under certain subjective limitations and hindrances, which, however, far from concealing it and making it unrecognizable, rather bring it out by contrast and make it shine forth all the more brightly. (*Gl*, 4:397.)

Kant then presents several examples for the sake of illuminating the distinction between merely acting in accord with duty and acting *from* duty (*Gl*, 4:397-9). All sorts of inclinations can incline a person to act in ways that happen to be in accord with duty. But

actions from duty issue solely from what Kant calls *respect for the law*, and it is this, says Kant, that endows them with moral worth (*Gl*, 4:400-1). Kant follows this, in turn, with the following:

Hence nothing other than the *representation of the law* in itself, *which can of course occur only in a rational being*, insofar as it and not the hoped-for effect is the determining ground of the will, can constitute the preeminent good we call moral, which is already present in the person himself who acts in accordance with this representation and need not wait upon the effect of his action.

But what kind of law can that be, the representation of which must determine the will, even without regard for the effect expected from it, in order for the will to be called good absolutely and without limitation? Since I have deprived the will of every impulse that could arise for it from obeying some law, nothing is left but the [universal conformity of actions with law as such], which alone is to serve the will as its principle, that is, *I ought never to act except in such a way that I could also will that my maxim should become a universal law.* (*Gl*, 4:402.)⁶

When the representation of the law is the determining ground of the will, this constitutes “that preeminent good we call moral”, says Kant. This could just be a reiteration of the claim that actions issuing from respect for the law have moral worth. But Kant appears also to be returning to the topic of what it is about a good will that makes it good “without limitation”. And his answer appears to be that the good will has as its principle the

⁶ My brackets in the above are to capture a correction to Gregor’s translation. She translates, “die allgemeine Gesetzmaessigkeit der Handlungen ueberhaupt” as “the conformity of actions as such with universal law” (Gregor, p.56). My disagreement regards Gregor’s placement of “as such”, which is to correspond to Kant’s use of “ueberhaupt”. I believe that Kant meant “ueberhaupt” to apply to “allgemeine Gesetzmaessigkeit”, rather than to “Handlungen” (actions). A further question may occur to the reader at this point. Should we translate “allgemeine Gesetzmaessigkeit” as “universal conformity with law” or as “conformity with universal law”? It seems that the former is a more accurate translation, but the latter fits better with what comes next, namely, the first formulation of the categorical imperative, which is concerned with universal laws (not explicitly with universal conformity with laws). If we take the former translation, we might wonder whether Kant is entitled to formulate the categorical imperative as he has. Perhaps Kant should be saying that the good will aims at *universal conformity with universal laws*: Possession of a good will is not just about aiming at conformity with universal laws *now and then*; it has to do with aiming to be *generally* in conformity with those laws.

universal conformity of its actions with law. The purpose or aim of a good will, then, is *its actions' conformity with (moral) laws*, and it is for this that all its means (e.g. talents of mind and qualities of temperament) are summoned.

A further clarification is perhaps in order at this point. As we have seen, Kant compares the good will to understanding, wit, judgment, courage, perseverance, etc., which we can regard as elements of a person's character. This suggests that Kant regards the good will as itself a character trait, a stable feature that an agent either has or lacks. (Thus, an agent who has a good will has a stable aim of conformity with the moral law, or, the (system of) universally valid practical principle(s).) And this is the reading I prefer. However, my reading is not entirely uncontroversial. Barbara Herman interprets a Kantian good will as something more episodic: an instance of willing properly, or, a *good willing*. What Herman means by a *good willing*, in turn, is an agent's making the motive of duty the only thing that determines some particular action of hers. And, according to Herman (p.16), it is good willing that Kant intended to illuminate with his discussion of the moral worth of actions from duty. I grant that, for Kant, actions from duty involve good willings, as Herman defines them. But I think it is a mistake to equate Kantian good wills with Hermanian good willings. As we have seen, Kant not only compares the good will to other character traits, he also says that the "distinctive constitution" of "the will which is to make use of these gifts of nature" is called "*character*" (*Gl*, 4:393). Moreover, I find Herman's interpretation difficult to square with Kant's claim that "weakness in the use of one's understanding coupled with the strength

of one's emotions is only a *lack of virtue* and, as it were, something childish and weak, which **can indeed coexist with the best will**" (*MdS*, 6:408; bold-type is mine). This claim seems to make sense only if we understand it as saying that even those (human beings) with the best *character* are vulnerable to episodes of weakness. For, if the good will is just a good willing (*a la* Herman), how can the *best will*, the *best willing*, occur when the agent is being "childish and weak"?

So, I understand the Kantian good will as a stable element of character which involves aiming at living in conformity with the moral law. The question I now want to ask is whether this tells us enough about the good will to explain why it has the undiminishable goodness and worth Kant attributes to it.

II. THE GOODNESS OF THE GOOD WILL

I think the key to understanding Kant on the goodness of the good will lies in understanding how the possession of a good will endows a rational agent with a certain intrinsic *aptitude* for living in conformity with the moral law. Now, Kant could be interpreted as explicitly rejecting *any* aptitude-based explanation of the goodness of a good will, when he says, "A good will is not good because of ... its fitness to attain some proposed end, but only because of its volition" (*Gl*, 4:394). What I will be suggesting here, however, is that it is precisely the good will's volition—its aiming at conduct in accord with the moral law (for its own sake)—that gives it a certain fitness for attaining not just any proposed end, but that of a lawful life. And I hope to show that this provides

a plausible explanation of the undiminishable goodness of the good will. This may be quite different from the explanation Kant would provide, but then again it could be Kant's. In any case, I think it is the best explanation available to him.

Perhaps the best way of getting at the sort of aptitude I have in mind is to start with what led me to think about it, something Kant says in the preface of the *Groundwork*:

For, in the case of what is to be morally good it is not enough that it *conform* with the moral law but it must also be done *for the sake of the moral law*; without this, that conformity is only very contingent and precarious, since a ground that is not moral will indeed now and then produce actions in conformity with the law, but it will also produce actions contrary to the law.⁷

Kant's point here seems to be that, unless an agent aims at conformity with the moral law, whether or not her action conforms with the law is largely a matter of accidents and/or coincidences. And this is supposed to explain why some actions that conform with the moral law are morally good, and why others are not. The moral goodness of actions in conformity with the moral law depends on the volitions behind them because their volitions determine whether (or the extent to which) their conformity is accidental or coincidental.

Although this is a point about what makes *actions* morally good, I will soon argue that a similar point can (and should) be made regarding the goodness of a good will. But let me first take a page or two to explain what I think is important and right about this explanation of the moral goodness of actions done "for the sake of the moral law". An action A of mine may be lawful, in the sense that it conforms with the moral law, even if

⁷ *Gl*, 4:390. Notice that Kant is here calling certain kinds of actions "morally good", and these are the very actions to which he elsewhere attributes *moral worth*.

my reasons for performing A had nothing to do with A's lawfulness. To put it another way, I can do the right thing without doing it for the right reasons. A clear example: I save a drowning child, but *only* because I expect to be admired and possibly rewarded for doing so. Many of the actions one might perform solely for the sake of admiration or reward happen to be lawful actions. But this is largely because many of the actions that *inspire* admiration and/or gratitude happen to be lawful actions. Thus, the actions that an admiration-seeker and/or a reward-seeker chooses will often turn out to be lawful actions even when lawfulness is not one of the criteria the agent applies in choosing her actions. The important point here, however, is that this is often *but not always* the case. In some situations—and they are hardly rare—what it takes for a person to gain another's admiration and what constitutes doing the right thing are very, very different. It depends on *whose* admiration is pursued.

We might be tempted to characterize this observation as one concerning the admiration-seeking volition's *reliability* in producing lawful actions: it is somewhat reliable, but not as reliable as a volition that is directed at lawful action itself. And we could plausibly say that the lawful-action-seeking volition is more reliable at producing lawful actions than most other volitions. We might even speculate that the lawful-action-seeking volition is more reliable (at least in our world as we know it) than *any other* possible volition. But even if this were correct, it is not quite enough to explain why lawful actions that *do not* issue from this volition *are not morally good at all*, which is Kant's position. For, if it is precisely a volition's reliability in producing lawful actions in

virtue of which a lawful action issuing from that volition is morally good, then it would seem that a lawful action issuing from any volition that has this reliability to *some* extent (e.g., the reward-seeking volition) is to some extent morally good. So, something else needs to be said about the lawful-action-seeking volition, something which is true of it but not at all true of any other volition, in order to make sense of the hard line Kant draws between the two kinds of lawful action.

Can we explain the (exclusive) moral goodness of actions “done for the sake of the moral law” by saying that those actions’ conformity with the moral law is *not at all* precarious or contingent—that actions issuing from the lawful-action-seeking volition *necessarily* conform with the moral law? If this were true it would give us one plausible way of explaining why these actions, and these alone, are morally good: The volition behind them is not just reliable, but *absolutely* reliable in producing lawful action, and this cannot be said of any other volition. The trouble, though, is that this volition *is not* absolutely reliable in producing action in conformity with the moral law. For, no matter how we (plausibly) conceive of the moral law, it is possible for an agent to act for the sake of it and yet fail to act in accord with it, simply through a failure in motor skills. (The action that is in fact performed is not the action the agent intended to perform. So, although she acted for the sake of the moral law, the action she performed was not in accord with the law.) But all is not lost, since we can easily amend the current suggestion by saying that what makes the lawful-action-seeking volition morally special is that,

absent failures in motor skills, the actions it produces are always in accord with the moral law, and this cannot be said of any other volition.

Now, arguably, there are further ways in which it is possible for an agent to act for the sake of the moral law and yet fail to act in accord with it. But the point I want to drive home here is that the volition involved in acting “for the sake of the moral law” (as Kant understands this) is, among possible volitions that rational agents might have as they act, especially well-suited for lawful conduct. For, when this volition *prevails* in determining an agent’s action, that action is necessarily lawful.⁸ Roughly: The agent correctly represents an action A as being in accord with the moral law, aims (because of this) to perform A, and succeeds; thus, it follows that the agent performs a lawful action. On the other hand, for almost any other possible volition an agent might have as she acts, when the volition prevails in determining an action, it is still “contingent and precarious” whether that action is lawful. Many of these other volitions can and often do produce actions that happen to be lawful, but since these volitions are not directed at lawful action itself, their prevailing does not guarantee it. This is what makes the volition in actions done “for the sake of the moral law” morally special. The actions that issue from it are not lawful due to accident or coincidence; rather, their lawfulness was the reason they were performed in the first place. And that, I contend, is a good Kantian way of explaining why these actions are morally good.

⁸ This way of understanding a volition’s relevance to its action’s moral goodness was inspired by Robert Johnson’s “Expressing a Good Will: Kant on Moral Worth”, *Southern Journal of Philosophy*, Summer, 34:2 (1996), pp.147-168.

Let us now recall that the good will, according to Kant, has as a stable and fundamental feature of her character a very similar volition: to *live* in conformity with the moral law. This volition, too, is morally special, in much the same way as the one just discussed. Among all possible volitions agents might have as stable and fundamental features of their characters, this one is especially well-suited for a lawful life. To be sure, since this volition is directed at a kind of *life*, rather than a kind of action, it takes a lot more for it to prevail. But when it does, the necessary product is a life in conformity with the moral law. On the other hand, for any volition that *does not* have living lawfully as its object, the volition's prevailing is no guarantee of a lawful life. For example, suppose I aim first and foremost at living a life of pleasure, and succeed at this. It may be the case that I thereby happen also to lead a lawful life if, say, the kinds of activities that give me the most pleasure happen to be in conformity with the moral law. On the other hand, it may instead be the case that I hardly ever am in conformity with the moral law if, for example, the activities I find most pleasant are not lawful. The lawfulness of a life successfully lived for the sake of pleasure is contingent on lawfulness coinciding with pleasure. And generally, the lawfulness of a life successfully lived for the sake of X is contingent on lawfulness coinciding with X. But if X is *living lawfully*, the coincidence is necessary. So, again, I agree with Kant that the goodness of the good will rests in its volition. But what makes the good will's volition morally special is that it is suited for lawful living in a way that other stable volitions are not.

Kant would surely agree, but perhaps only to a point; he may think that the goodness of a good will has to do with more than just the fact that its volition has this fitness for lawful living. He might want to say, instead or in addition, that the good will's goodness is essentially tied to its "inner freedom". Kant famously posits the idea that human beings can choose—in a way that is independent of natural causes—to adopt the aim of living in conformity with the moral law. But if the good will's goodness rests (at all) in its being the product of free choice, on Kant's view, this is one junction at which he and I part company. We can only speculate on whether human beings are capable of supernatural choice, and I do not think we need to presuppose it in order to arrive at a plausible account of the good will's goodness. (I will discuss this issue at length in Chapter Two.) In any case, my aim here is to see how good of a conception of moral virtue we can get without it.

So far I have argued that the good will is morally special because there is a necessary connection between the prevailing of its volition and living in conformity with the moral law. But I should stress that an agent's possession of a good will is, even on Kant's view, no guarantee that this volition *will prevail*. Even if I aim first and foremost at living a lawful life, this does not mean that I can overcome all temptations to violate the moral law—not even all *foreseeable* temptations. An agent who has a good will need not also have a strong will. Moreover, I may have to cultivate certain "talents of mind", "qualities of temperament", and other capacities, in order to live lawfully. In short, I may need to augment my aptitude for lawful conduct, if I am to live lawfully. In the next section, I

present what I call the five perfections, elements of character that agents have to varying extents, and which (I argue) determine their aptitude for lawful conduct. One of these is the stable volition to live lawfully for its own sake; I call this the moral commitment, and I call the agent who has it morally committed. The other four perfections are intended to exhaust the elements of character needed to make sure that this volition prevails. I define the ideally virtuous agent as the agent who has each of these five perfections to the fullest extent. By definition, then, her moral commitment necessarily prevails, if my account of the perfections is right. And, my suggestion is, the rest of us should be measured by the extent to which we approximate her, along each of five corresponding dimensions.

III. APTITUDE FOR LAWFUL CONDUCT

An agent has the moral commitment if she has a stable, fundamental aim of living in accord with whatever practical principles are valid for all rational beings (though she might not be clear on what those principles are). I call this a *fundamental* aim to indicate only that the aim is not entirely parasitic on her other aims, desires, or whatever; in other words, the agent has not adopted this aim just as a means for the achievement of other aims she has. So, the agent who has this commitment does not aim at living a moral life *just* because she expects this will make her happy, for example, nor *just* because she expects others to respect her for it. She may have these expectations, and they may be a part of why she wants to live a moral life. But she only has the moral commitment if she also thinks of the moral life as worth striving for in its own right. And, for the *ideally*

virtuous agent, living lawfully is the *most* important and valuable thing she strives for, and she holds everything else (even the sum of all other things) to be subordinate to it. This is what I will call having the *perfect* moral commitment. I should emphasize that, although it is plausible that many of us have the moral commitment, as I have described it, I would guess that few of us have the *perfect* moral commitment. That is, although many of us find it important and valuable to live a moral life, few of us take this to be more important and valuable than all other things combined.⁹

I should point out that the moral commitment, as I have defined it, also has the special fitness we attributed to the good will in the previous section. (And I am not just talking about the *perfect* moral commitment.) We saw that there is a necessary connection between the prevailing of the good will's volition and lawful living. The same necessary connection holds between the prevailing of the moral commitment and lawful living: If I aim to live in accord with whatever practical principles are valid for all rational beings, and succeed in this, it follows that I live lawfully.

Of course, there is nothing inconsistent in saying that agents who have the perfect moral commitment can be less than ideally virtuous. One way in which they can fall short of the ideal is by lacking the ability to overcome episodic temptations to deviate from doing what they believe is morally best. In general, I will think of weakness of will as the vulnerability to episodes in which one gives in to temptations to do things that one

⁹ On my reading of Kant, the perfect moral commitment is one of the distinctions of the good will. But, in the *Religion*, Kant seems to say that one either has the perfect moral commitment or one is not at all morally committed. In the next chapter, I argue against this idea, and discuss agents who are morally committed but not perfectly morally committed.

sees as conflicting with the aims one values most (*whatever* these are), or as impairing the pursuit of those aims. Strength of will is the capacity to overcome such temptations. I suspect that we all have this capacity to some extent but most, if not all, human beings give in to temptation from time to time. They do not have *perfect* strength of will: the *unfailing* ability to overcome *any* temptation to do what conflicts with the aims one values most.

If an agent has the perfect moral commitment, the aim she values above all others is living lawfully. And if, in addition, she has perfect strength of will, she will always overcome temptations to do things that are inconsistent with that aim. But this still does not mean that she will always act lawfully. For, if she has the wrong ideas about the moral law, acting on those ideas could fail to be lawful. So, having inadequate knowledge of the moral law is another way of falling short of the ideal. The ideally virtuous agent has a complete and correct account of the moral law, and always can immediately tell whether an action is or is not lawful. This is what I mean by the ideal of perfect moral understanding.¹⁰

But even if an agent has each of the above three perfections, she can still fail to do what is lawful if she does not know enough about *her world*. For example, even if she tries with all her might to benefit others, she may go about it the wrong way and wind up

¹⁰ Texts in the *Groundwork* suggest that Kant was overly optimistic about our ability to acquire perfect moral understanding, though Kant seems less optimistic in *The Metaphysics of Morals*, where he includes cultivating one's moral understanding as a component of the duty of self-perfection. Kant also seems, in the *Groundwork*, to presuppose that the good will has perfect moral understanding as well as the perfect moral commitment.

harming them instead. A doctor who aims at promoting her patient's health may not know enough about how to interpret his symptoms and may, as a result, wind up administering the wrong treatment. In an attempt to comfort a grieving friend, a person might say something that strikes the wrong nerve and upsets her even more. Motivated by environmentalist concerns, an engineer might design an automobile engine that runs cleaner but, unbeknownst to him, can only be mass-produced through processes that are extremely harmful to the environment. My point here is just that people can fail in their sincerest and most admirable efforts to do what is right if their understanding of the way things work in our world is incorrect or incomplete. What is lacking in such cases is empirical understanding. Now, it might be too much to say that the ideally virtuous agent knows everything there is to know about her world. (Why, for example, would she have to know at all times the exact position of every subatomic particle in the world?) But I want to at least define her as knowing every empirical fact that is potentially relevant to the lawfulness of her available actions. And this is what I will have in mind when I say that she has perfect empirical understanding.

Finally, an agent may have all four of the above perfections and yet fail to do what is lawful simply because she is not fully capable of doing what she aims to do. A surgeon who knows perfectly well what sort of procedure is best for a patient may lack the dexterity to successfully execute it. A fireman who tries to rescue several people from a burning building may lack the strength, speed, and/or endurance to save them all. To be sure, neither person is obviously culpable. But we can plausibly say that each fails to do

something that would have been lawful, and that this is due to a lack of empirical power. *Perfect* empirical power consists in having the capacity to do anything that is compatible with the physical laws.

This is quite an ideal, but the point of discussing it is to highlight what I think are the key dimensions of aptitude for lawful conduct. I submit that the ideally virtuous agent, as I have described her, never fails to act lawfully. She knows the moral law, wants above all else to live in accord with it, knows in all cases what available actions accord with it, and never lacks the know-how, physical capacity, or strength of will to perform them. Our deviations from lawfulness are explainable by the fact that we fall short with respect to one or more of the five perfections just described.

This brings us to another sense in which the moral commitment is especially suited for lawful conduct. We already have seen that it is special in the sense that, when it prevails, a lawful life is the necessary result. But, for a morally committed agent, the rational response to the recognition that one can improve one's strength of will, moral understanding, empirical understanding, or empirical power, is to aim at that improvement. If I am at all committed to living lawfully and know that I can make myself better suited for it, this gives me a reason to aim to do so. And insofar as I prevail in *this* aim, the extent to which my moral commitment prevails becomes less and less a matter of coincidence. (I will develop this point further in Chapter Four, where I argue for a duty of moral self-perfection.)

These results should hardly surprise us. From the fact that I have a volition that is directed at X, it follows that I achieve X when this volition prevails. It also follows that the rational response for me, when I recognize that I can improve my aptitude for achieving X, is to aim to do so (provided that this is worth the opportunity costs to me). This is so whether X is *being a good carpenter, becoming a millionaire, writing a best seller*, or whatever. As regards the first claim, surely one can become, say, a good carpenter by aiming at something other than becoming a good carpenter. If I aim, first and foremost to make a comfortable living, and go into business as a carpenter, the market and my responses to it may make it the case that I become a good carpenter in the course of making a comfortable living—even if being a good carpenter itself never becomes something I value in its own right and aim at for its own sake. But this would not work out if there were no demand whatsoever for carpentry. The prevailing of the living-comfortably volition does not guarantee good carpentry; the only volition that guarantees this when it prevails is a good-carpentry volition. As regards the second claim, *whether* the good-carpentry volition prevails depends on the resources the agent has available to her, including her knowledge of what makes carpentry *good* carpentry, her knowledge of the techniques and procedures that produce it, and her ability to execute them. And so, depending on *how* committed the agent is to being a good carpenter, her recognition that she can improve her knowledge and/or ability to execute rationally commits her to improving her aptitude for good carpentry. This is just a fact about rational agents' seeing the point of improving themselves for the sake of their pursuits.

Most accounts of virtue have it that the virtuous agent does the right thing for the right reasons. This view takes this idea very seriously, but applies it in a very general way. An agent may perform an overt action that is the right one to perform (in the circumstances), and may do so because that action is honest, beneficent, etc. She aims at honesty, beneficence, etc., and this is her reason for performing the action. But the account offered here also asks whether she has the right reasons for *adopting and maintaining* the aims of beneficence, honesty, etc., which we might call right *inner* actions. We might be tempted to say that a person is, at least to some extent, virtuous as long as she has the right aims (so that she “has the right reasons” for performing the right *overt* actions she performs). But, at the very least, I would want to add that the agent who *adopts and maintains the right aims* for the right reasons (or, who performs the right *inner* actions for the right reasons) is more virtuous than one who simply has acquired the right aims, or who adopts and maintains them for the wrong reasons. I am most sympathetic, however, to a stronger view that takes possession of the moral commitment as a necessary condition for virtue. (I will offer an argument for this view in Chapter Three.) Thus, my view would have it that a person can be morally good, in the sense that she has the right aims, even if she is not virtuous, which requires having those aims for the right reasons. Like Kant, I see the right reasons regarding these inner actions (adopting or abdicating aims) as being provided by the criterion of lawfulness: universal validity for rational beings. And the morally committed agent *takes* considerations of lawfulness to provide reasons for her to adopt or abdicate aims.

Thus, this view has it that, in morally appraising agents (ourselves and others), we should assess the extent to which we/they have each of the five perfections, with pride of place going to the extent to which the agent is morally committed. Again, even the perfect moral commitment (an agent's being committed to living lawfully above all else) does not guarantee a lawful life, but the lack of the moral commitment makes the lawfulness of one's conduct much more contingent on things other than one's character. And if we are morally appraising character, it makes sense to regard a feature of character as superior if makes lawful conduct less of an accident than other features an agent might have in its place. I hope to have shown that, as aims go, the aim of living lawfully for its own sake is, in this regard, better than any other aim an agent might have in its place.

Again, it matters whether the morally committed agent has the right account of the moral law, and knows how to live her life in conformity with it. Even an agent who is committed first and foremost to living in accord with whatever practical principles are valid for all rational beings can violate the moral law (and perform actions that are not only wrong but have terrible consequences) if the principles she regards as universally valid are in fact not, or if she has misunderstood the implications they have for her. But this is not a strike against the idea that the aim of lawfulness (for its own sake) is the morally superior aim. Ignorance and faulty reasoning can prevent one from living lawfully *no matter what one's aims are*. But if one does not aim at living lawfully itself, one can fail to live lawfully even in the absence of ignorance and faulty reasoning. Moreover, the moral commitment provides an incentive to *guard against* ignorance and

faulty reasoning, that is, to acquire moral understanding. (If I aim first and foremost to live in conformity with whatever practical principles are valid for all rational beings, I will want to know what those principles are and how a being in my circumstances can accord with them.)

So, although the extent of an agent's moral understanding should be regarded as relevant to a moral appraisal of her character, on this view, a deficiency with regard to moral understanding is not enough to diminish the moral goodness an agent has in virtue of having the moral commitment—even if *moral misunderstanding* leads her to commit terrible acts. Rather, an agent's moral understanding is seen as something that *adds to* the moral goodness of the moral commitment, by making it to an even lesser extent a matter of coincidence whether the agent lives lawfully. Similarly, this view sees the morally committed agent's empirical understanding, empirical power, and strength of will as adding to the moral goodness of the moral commitment. But none of these other four perfections can, on this view, add any moral goodness to an agent that altogether lacks the moral commitment. For, even if she has each of these four perfections in its ideal form, the lawfulness of her conduct is as much a matter of coincidence and accident as it would be if she had each to only a minimal degree.

Accordingly, this view has it that only morally committed agents truly admit of any degree of virtue, and that *how virtuous an agent is* depends entirely on the extent to which she has each of the five perfections.

CHAPTER TWO

NATURAL MORAL GOODNESS

INTRODUCTION

In the previous chapter, I argued that the moral appraisal of an agent should focus on her aptitude for lawful conduct, as determined by the extent to which she has each of the five perfections: the moral commitment, moral understanding, strength of will, empirical power, and empirical understanding. I equated a person's moral goodness with her aptitude for lawful conduct, and suggested reasons for thinking of the moral commitment (an agent's having living in accord with the moral law as a stable, fundamental aim) as the cornerstone of an agent's moral goodness. Although this view is largely inspired by ideas Kant presents in *Groundwork of the Metaphysics of Morals*, some of the texts in his *Religion within the Boundaries of Reason Alone* suggest that his view differs from mine in at least two respects. First, Kant argues that an agent cannot be called "morally good" unless her choice to live in accord with the moral law is *free*, that is, independent of natural causes (*Rel*, 6:21). (Call this *the freedom thesis*.) The view I am proposing has it that an agent's moral goodness does not depend on how she comes to aim at living lawfully; her having the moral commitment is sufficient for her being a morally good agent. Second, Kant argues that an agent is not morally good unless she has living in accord with the moral law as her *sole* stable and fundamental aim (*Rel*, 6: 22-5). (Call

this *the exclusivity thesis*.) My view has it that an agent can be morally good even if she has, in addition, *other* stable and fundamental aims.

At first glance, both the freedom thesis and the exclusivity thesis appear to have some unappealing implications. If human beings are not capable of free choice, it follows from the freedom thesis that no human being can be properly called “morally good”. But, one might object, even if all human choices were entirely due to natural causes, it would still make sense to say that some human beings are nevertheless morally good. If, for example, a human being has developed, through natural causes alone, into a person who reflects on the way she lives, reasons correctly about how she *ought* to live, and goes to great lengths to organize her life accordingly, it seems implausible to deny that she is a morally good human being. The problem with the exclusivity thesis, on the other hand, is that it seems to set too high of a bar for moral goodness. It says that any human being who has stable and fundamental aims in addition to that of accord with the moral law is not morally good *at all*. But this seems to imply that such a person is, from a moral perspective, no better than a person who does not care about living in accord with the moral law, which is also implausible.

In this chapter, I examine Kant’s arguments for these two theses. I then argue first that, once we understand what Kant means by ‘morally good’ (in the relevant texts of the *Religion*), we can see that the above objection against the freedom thesis misses the point Kant was trying to make, and that this point is not in conflict with the idea that moral goodness should be construed solely in terms of aptitude for lawful conduct. Second,

however, I point out that Kant lacks an argument for the exclusivity thesis, and argue that it can plausibly be rejected.

I. THE FREEDOM THESIS

Whether an agent is morally good or evil depends, according to Kant, on the practical rule, or, maxim, that she sets for herself as *the* rule that will govern *all* of her actions, or, her *fundamental* maxim. Kant's so-called "incorporation thesis" says that agents freely choose this maxim:

[F]reedom of the power of choice has the characteristic, entirely peculiar to it, that it cannot be determined to action through any incentive *except so far as the human being has incorporated it into his maxim* (has made it into a universal rule for himself, according to which he wills [sic?] to conduct himself); only in this way can an incentive, whatever it may be, coexist with the absolute spontaneity of the power of choice (of freedom). (*Rel*, 6:23-4.)

As we will see, Kant asserts that if the *incorporated* incentive is the moral law itself, the agent is morally good; otherwise, the agent is evil (*Rel*, 6:24). The morally good agent makes the moral law into a universal rule for herself; it is her fundamental maxim, and all of the more particular practical rules on which she acts, her particular maxims, are derived from it. So, we can see that the agent who has the moral law as her universal rule is just what I call the perfectly morally committed agent. That Kant is committed to the exclusivity thesis (only the agent who is *perfectly* morally committed is morally good) follows from his assertion that any agent who does not make the moral law her *universal* rule is evil. His arguments for this will be the subject of the next section.

But the above passage also suggests that Kant endorses the freedom thesis: one is not morally good unless one's moral commitment itself is chosen in a way that is independent of natural causes. For, *incorporating an incentive into one's maxim* involves not only making that incentive into a universal rule for oneself, but also doing this through "the absolute spontaneity of the power of choice (of freedom)". And a passage earlier in the *Religion* both (a) confirms that this spontaneity or freedom involves choice that is independent of natural causes, on Kant's view, and also (b) explains *why* Kant posits this special (supernatural) kind of choice in the first place:

[The] subjective ground ... of the exercise of the human being's freedom in general ... antecedent to every deed that falls within the scope of the senses ... must itself always be a deed of freedom (**for otherwise the use or abuse of the human being's power of choice with respect to the moral law could not be imputed to him, nor could the good or evil in him be called "moral"**). Hence the ground of evil cannot lie in any object *determining* the power of choice through inclination, not in any natural impulses, but only in a rule that the power of choice itself produces for the exercise of its freedom, i.e., in a maxim. One cannot, however, go on asking what, in a human being, might be the subjective ground of the adoption of this maxim rather than its opposite. **For if this ground were ultimately no longer itself a maxim, but merely a natural impulse, the ... exercise of freedom could be [entirely] traced back to a determination through natural causes—and this would contradict freedom.** (*Rel*, 6:21; bold-type is mine.)

An exercise of freedom cannot be entirely "traced back" to natural causes, says Kant, and it is only deeds of freedom that are imputable to agents. Thus, if the "good or evil" in a human being were entirely due to natural causes, it could not be properly imputed to her, and so could not be called "moral". So, in order for a human being to be *morally* good or evil, according to Kant, her choice with respect to the moral law cannot be merely an

effect of natural causes; it must be “a deed of freedom”. Even if an agent chooses the moral law as her universal practical principle, she is not morally good if that choice is entirely an effect of natural causes.¹¹

This shows that Kant’s commitment to the freedom thesis is based on a connection he sees between imputability and the propriety of calling the good or evil in an agent “moral”. Now, in the introduction to this chapter, I presented an objection to the freedom thesis, and asserted that it misses the point Kant was trying to make. The objection was as follows: If a human being has developed, through natural causes alone, into a person who reflects on the way she lives, reasons correctly about how she *ought* to live, and goes to great lengths to organize her life accordingly, it seems implausible to deny that she is a morally good human being. Kant would say that the good in this agent cannot be called “moral”, since the good in her is entirely due to natural causes, and so not imputable to her. But this would hardly satisfy the objector, since it does not explain why there can be no such thing as *wholly natural* moral goodness. Surely, the objector might add, an agent’s freely choosing to live in accord with the moral law does not make her moral commitment more effective in determining lawful conduct than any naturally caused moral commitment. In what sense, then, is a freely chosen moral commitment better,

¹¹ Notice, what Kant is saying here is not merely that the agent must *see* “this ground” as being something that could not be entirely “traced back to a determination through natural causes” (in order for the ground to be “a deed of freedom”). If the subjective ground is in fact traceable back to a determination through natural causes, this by itself is sufficient for it *not* to be a deed of freedom—even if the agent herself thinks it *is* a deed of freedom.

from a moral perspective, than a naturally caused one? It is questions like these that miss the point Kant is trying to make.

The idea in the above passage is that whether the good (or evil) in an agent can be called “moral” depends on whether it is imputable to her, rather than entirely traceable to natural causes. Let us now note that one paragraph later in the *Religion*, Kant asserts that “nature is not to blame for it (if the character is evil), nor does it deserve praise (if it is good), but that the human being is alone its author” (*Rel*, 6:21-2). The idea there is that whether an agent deserves praise (or blame) for her character depends on whether she, rather than nature, authors her character. For Kant, then, if an agent’s good (or evil) character is entirely a product of natural causes, it could not be called “moral”, nor would the agent deserve praise (or blame) for it. Two points are worth emphasizing here. First, Kant has tacitly made room for the possibility of good or evil character that *cannot* be called “moral” and whose possessor deserves neither praise nor blame for having it. Second, the deciding factor for Kant is whether the good or evil character is imputable to its possessor or, instead, entirely due to natural causes. If my character is *naturally* good or evil, it is not imputable to me; so my character cannot be called “moral” and I am not to be praised or blamed for it. On the other hand, if I have *freely chosen* to be good or evil, my character is imputable to me, in which case it *can* be called “moral” and I can properly be praised (for being good) or blamed (for being evil). Kant evidently thinks that calling an agent’s good or evil character “moral” is the same as saying that she can properly be praised or blamed for it. The point Kant is trying to make (at least in *Rel*

6:21), then, is not that a freely chosen moral commitment is better, from a moral perspective, than a naturally caused one. It is just that an agent cannot properly be praised for the good in her (or blamed for the evil in her) if it is just an effect of natural causes.

So, Kant is simply drawing a distinction between: (a) whether a person has (or lacks) the moral law as a universal rule—in which case he would call her good (or evil) *simpliciter*—and (b) whether her having (or lacking) this universal rule can be imputed to her—in which case he would say that the good (or evil) in her can be called “moral”. Understood in this way, Kant’s freedom thesis really is not at odds with thinking of moral goodness in terms of aptitude for lawful conduct. The apparent conflict is due only to Kant’s terminology in the above. He simply uses ‘good’ where I would use ‘morally good’, and his use of ‘moral’ is intended to connote imputability.

Of course, this is not to say that the agents whom I would call “morally good” are precisely those whom Kant would call “good”. For Kant, it is only those agents who have the moral law as a *universal* rule that are “good”; all others are evil. In other words, Kant is committed to the exclusivity thesis, to which we now turn.

II. THE EXCLUSIVITY THESIS

I have already mentioned in passing the exclusivity thesis’ connection to the “incorporation thesis”. To understand Kant’s motivation for each of these, let us now look at the more of the passage in which he presents the latter. The key paragraphs

(which I label Pa1, Pa2, and Pa3, respectively) are presented in their entirety and with nothing in between omitted (*Rel*, 6:22-5):

It will readily occur to anyone to ask ... whether some might not claim that the human being is by nature neither of the two [neither good nor evil], others, that he is both at once, that is, good in some parts and evil in others. Experience even seems to confirm this middle position between the two extremes.

[Pa1:] It is of great consequence to ethics in general, however, to preclude, so far as possible, anything morally intermediate, either in actions (*adialora*) or in human characters; for with any such ambiguity all maxims run the risk of losing their determination and stability. Those who adhere to this strict way of thinking are commonly called *rigorists* (a name intended to carry reproach, but in fact a praise); so we can call *latitudinarians* those at the opposite extreme. These latter, again, are either latitudinarians of neutrality and may be called *indifferentists*, or latitudinarians of coalition and can then be called *syncretists*.

[Pa2:] On the rigorist's criteria, the answer to the question just posed is based on the morally important observation that freedom of the power of choice has the characteristic, entirely peculiar to it, that it cannot be determined to action through any incentive *except so far as the human being has incorporated it into his maxim* (has made it into a universal rule for himself, according to which he wills [sic?] to conduct himself); only in this way can an incentive, whatever it may be, coexist with the absolute spontaneity of the power of choice (of freedom). But the moral law is itself an incentive in the judgment of reason, and whoever makes it his maxim is *morally* good. Now, if the law fails nevertheless to determine somebody's free power of choice with respect to an action relating to it, an incentive opposed to it must have influence on the power of choice of the human being in question; and since, by hypothesis, this can only happen because this human being incorporates the incentive (and consequently also the deviation from the moral law) into his maxim (in which case he is an evil human being), it follows that his disposition as regards the moral law is never indifferent (never neither good nor bad).

[Pa3:] Nor can a human being be morally good in some parts, and at the same time evil in others. For if he is good in one part, he has incorporated the moral law into his maxim. And were he, therefore, to be evil in some other part, since the moral law of compliance with duty in general is a single one and universal, the maxim relating to it would be universal yet particular at the same time: which is contradictory.

The question Kant aims to answer here is whether it is possible for a human being to be *neither* good nor evil (as “indifferentists” think) or for a human being to be *both* (as “syncretists” think). Kant aligns himself with “rigorists”, who reject both of these intermediary possibilities. And Pa1 suggests that at least part of his reason for doing so is his thinking that the respective views of the indifferentist and the syncretist involve some unacceptable ambiguity.

In Pa2, Kant argues against the indifferentist position, by appealing to the rigorist’s observation (the “incorporation thesis”) that the only way for an incentive to determine an action in a way that is compatible with the “absolute spontaneity of the power of choice” is through the agent’s *incorporating that incentive into her maxim*: her making the incentive into a universal rule for herself, thereby aiming to generally conduct herself in accord with that rule. An agent who makes the moral law into such a universal rule for herself is, says Kant, morally good.¹² On the other hand, if an agent freely chooses some action A when she knows both that A is not in accord with the moral law and that some other available action B *is* lawful, this (according to Kant) means that she has incorporated *deviation from the moral law* into her maxim. She *does not* aim at *generally* according with the moral law, since she permits herself to deviate from it. And

¹² Notice, making some putative rule a universal rule for myself must be distinguished from making some maxim a rule for myself because *it is universalizable*. The former involves making something into a rule that I will *universally follow*, or, *universally accord with*—in the sense that I allow myself no exceptions to this rule—which is different from my seeing a rule as universally valid, that is, as valid for all rational beings. I point this out only because it would be easy to assume that whenever Kant uses ‘universal’, he has the latter sense in mind. But this is not what he has in mind here.

this, says Kant, makes her “an evil human being”. Thus, Kant concludes (against the indifferentist), it is not the case that human beings are neither good nor evil.

In Pa3, Kant again appeals to the idea that morally good agents aim at *generally* acting in accord with the moral law, this time to argue against the syncretist position that a human being can be *both* morally good and evil (at the same time). The precise content of this argument is difficult to discern. But the idea seems to be that there is something contradictory in supposing that an agent could be morally good in some parts but not in all parts, since an agent’s moral goodness hinges on her making the moral law a *universal* rule for herself. One way of making sense of this is to say that if I have some stable and fundamental *non-moral* aims (e.g., wealth and political power) in addition to that of accord with the moral law, and I pursue each of these aims, the moral law is not my *general* (or, *universal*) rule, since it is not the rule that generally determines my actions. For, I also have set for myself certain *other* practical rules (e.g., Do what promotes my wealth/political power) that are independent of the rule: Act in conformity with the moral law. And so if it is sometimes one of these other practical rules that I act on, and not the rule corresponding to the moral law, then the moral law is not a universal rule for me. And, if moral goodness attaches only to those agents for whom the moral law is a *universal* rule, as Kant claims, I am not morally good.

Recall, the exclusivity thesis has it that only those agents who are perfectly morally committed are morally good. In other words, moral goodness attaches only to those agents who have conformity with the moral law as their *sole* stable and fundamental aim.

In Kantian terms, only those who have made the moral law into a *universal* rule for themselves are morally good. But Kant does not explicitly argue for the exclusivity thesis. He *has*, in effect, shown that one cannot have the moral law as a universal rule while also having another rule that allows one to deviate from the moral law under certain circumstances. And this is correct. (For any putative rule R, it is impossible for me to make R a universal rule for myself while allowing exceptions to it.) But, strictly speaking, it does not follow from this that only those who have the moral law as a universal rule are *morally good*. The exclusivity thesis is a substantive claim about the moral goodness of agents, a claim which Kant simply presupposes in his arguments against the indifferentist and the syncretist. What I want to propose next is that it is more plausible to say instead that *any* agent who has the moral commitment to *any* extent is in virtue of this to some extent morally good. In Kantian terms, any agent who makes the moral law into a rule for herself is morally good (even if she does not make it a *universal* rule for herself).

III. AGAINST THE EXCLUSIVITY THESIS

Imagine an agent A, who finds it important in its own right to live in accord with the moral law, but who also finds it somewhat important to lead a pleasant life. Suppose that in almost all cases where she would have to choose between doing something that accords with the moral law and doing what is most pleasant, she would choose the lawful path; but not in all cases, since *as a rule* there is only so much pleasure she is willing to

forego for the sake of doing what is in accord with *all* of the practical principles she regards as valid for all rational beings. (She may, for example, think that one should not lie, but knows she would lie if the payoff were huge and “nobody would get hurt”).¹³ Now compare A to an agent B, who does not care at all about living lawfully, and basically does whatever she can get away with that best suits her mood. According to the exclusivity thesis, neither A nor B admits of any moral goodness. This result seems wrong. And the previous chapter gives us a way of explaining what is wrong with it. B has *no* aptitude for lawful conduct. Whether she does what is lawful is, on any occasion, entirely a matter of coincidences between what she happens to feel like doing, what she can get away with, and what is lawful. An action’s lawfulness (or its *apparent* lawfulness to her) would never be among B’s reasons for performing it. And in this B differs from an agent C who has only the most minimal moral commitment.

¹³ This may seem to be a misguided example, since it seems to contradict the idea that, for Kant, the moral law just is Kant’s categorical imperative: “I ought never to act except in such a way that I could also will that my maxim should become a universal law” (*Gl*, 4:402). But, first, we could simply sidestep this worry by assuming that A has misunderstood what the moral law is. Second, Kant seems to say, in the preface of the *Groundwork*, that the prohibition on lying is itself a moral law:

Everyone must grant that a law, if it is to hold morally, that is, as a ground of an obligation, must carry with it absolute necessity; that, for example, the command “thou shalt not lie” does not hold only for human beings, as if other rational beings did not have to heed it, and so with all other moral laws properly so called; that, therefore, the ground of obligation here must not be sought in the nature of the human being or in the circumstances of the world in which he is placed, but a priori simply in concepts of pure reason; and that any other precept, which is based on principles of mere experience—even if it is universal in a certain respect—insofar as it rests in the least part on empirical grounds, perhaps only in terms of a motive, can indeed be called a practical rule but never a moral law. (*Gl*, 4:389.)

As we have seen, Kant not only would have to say that A, B, and C are not morally good; he is also committed to saying that each of these agents is morally *bad*. Moreover, Kant appears to lack reasons for saying that any one of these agents is morally worse than the others. Strictly speaking, these two observations cannot be counted as strikes against the exclusivity thesis, since it is concerned only with moral goodness (and not at all with moral badness or evil). I bring them up, however, to highlight an advantage of thinking of the moral goodness or moral badness of agents in terms of their aptitude for lawful or unlawful conduct, where conduct is unlawful to the extent that it violates any of the practical principles that are valid for all rational beings. Notice, although B (who, recall, basically does whatever she feels like doing) lacks any aptitude for lawful conduct, she also lacks any aptitude for unlawful conduct. For, her engaging in unlawful conduct is as much a matter of coincidence as her engaging in lawful conduct. She aims at neither, and nothing in my description of her suggests that she is more likely to engage in one than the other. So, why call her morally bad? Why not say instead that she is *neither* morally good *nor* morally bad? Compare B to D, an agent who aims first and foremost at living unlawfully for its own sake, and who strives to make herself *better suited* for violating the moral law. It seems undeniable that D is morally bad in a way that B is not. And Kant's way of distinguishing morally good agents from morally bad agents appears incapable of accounting for this.

I submit therefore that it makes more sense to say that any agent who has the moral commitment to any extent is in virtue of this to some extent morally good. And that any

agent who has the *immoral* commitment (that is, who has some stable and fundamental aim of living an unlawful life) is in virtue of this to some extent morally bad. And any agent who has neither commitment to any extent is therefore neither morally good nor morally bad.¹⁴

In the previous chapter, I proposed that we should think of agents' moral goodness entirely in terms of their aptitude for lawful conduct, where this is understood in terms of the extent to which they possess the five perfections: the moral commitment, moral understanding, strength of will, empirical power, and empirical understanding. And I hope it is clear now what sort of account of moral *badness* (or, evil) I would propose. An agent's moral badness would be a matter of her aptitude for unlawful conduct. The cornerstone of aptitude for unlawful conduct would be the immoral commitment. But, interestingly, the immorally committed agent's aptitude for unlawful conduct is also a function of the extent to which she has the other four elements of character I have described as *perfections*: moral understanding, strength of will, empirical power, and empirical understanding. If I aim to live unlawfully, it "helps" that I know the moral law and what it would take to live in accord with it, since this knowledge tells me what sorts of conduct to avoid, and thereby makes it less a matter of coincidence whether I succeed in my aim to violate the moral law. Similarly, I need strength of will in order to make it the case that I can prevail in my aim to live unlawfully despite my "pesky" inclinations to be, say, kind or honest. And similar things hold for empirical power and empirical

¹⁴ Thus, we could say at least that the indifferentist is not completely wrong.

understanding. So, just as Kant says that “talents of mind” and “qualities of temperament” can be “extremely evil and harmful if the will which is to make use of [them] is not good” (*Gl*, 4:393), one could plausibly say that these four perfections are only desirable from a moral perspective on the condition that the agent who has them is also morally committed. This is just one more reason for seeing the moral commitment as the cornerstone of virtue and moral goodness.

CHAPTER THREE

VIRTUE AND PRACTICAL LAWS

INTRODUCTION

Everyone must grant that a law, if it is to hold morally, that is, as a ground of an obligation, must carry with it absolute necessity; that, for example, the command “thou shalt not lie” does not hold only for human beings, as if other rational beings did not have to heed it ... (*Gl*, 4:389.)

So far, I have been describing the moral commitment in a number of different ways. Mostly, I have called it the stable, fundamental aim of living in accord with universally valid practical principles. But I’ve also referred to the object of this aim as *living in accord with the moral law*, *living lawfully*, and *living a moral life*. At some points I simply matched the terminology of the Kantian texts I was working from; other times, I was deliberately vague because my provisional accounts of virtue and moral goodness were intended to be compatible with any plausible moral theory. In this chapter, I will make things more explicit and move toward a more distinctive account of virtue. The refinements to be defended here are inspired by one of the ideas expressed in the above excerpt, namely, that the most fundamental moral requirements are valid not just for human agents, but for all rational beings. I will now define the moral commitment as the stable, non-instrumental aim of living in accord with the practical principles that are valid for *all possible* agents. For brevity, I will follow Kant in calling these principles *practical laws*. And my primary goal here is to further defend an idea I hinted at in

Chapter One, namely, that the moral commitment should be regarded as necessary for virtue.

I will begin, however, by arguing that human agents who want to approximate perfect moral understanding should be concerned with identifying practical laws. I'll start by briefly defending and then employing the Virtue-Ethical canon that the virtuous agent does the right things for the right reasons. The key point there will be that when we apply this canon to the virtuous agent's moral *thinking*, we get the result that she grounds her beliefs about how she should live on her beliefs about practical laws. In this sense she thinks like a Kantian and, I contend, so should any human agent who sincerely aims to approximate virtue. In the second part of this chapter, I turn to the question of why we should see the moral commitment as necessary for virtue. I will show that if our account of virtue omits this requirement, then even the *ideally* virtuous agent's accord with moral requirements is to a great extent dependent on things that are external to her character. My account, on the other hand, enables us to maintain that the ideally virtuous agent *always* does what is morally best—whether she lives in this world or any other.

I. MORAL UNDERSTANDING AND PRACTICAL LAWS

Since I will be making use of the idea that the virtuous agent always does the right things for the right reasons, let me start by saying why I think we should accept this idea in the first place. I do not think we should just take this for granted. We have to admit that most thought about what an ideally virtuous human being would be like is very much

speculative. I cannot claim to ever have known, met, or even heard of an actual exemplar, and I doubt that I know anyone who can. Even if we focused just on the moral commitment, moral understanding, and strength of will, I'm pretty sure that none of us can seriously claim to know someone who has achieved each of these in its fullest form. In fact, all we can reasonably say is that some human agents have made exceptional progress. And how much evidence they might provide about the ideally virtuous agent is really not clear, since we cannot tell how close they are. So why, exactly, should we think that she does the right thing for the right reasons, given that we cannot confirm it empirically?

When "doing the right thing" refers to an overt action, my answer starts again with the Kantian idea (exploited in Chapter One) that when an action *was not* done for the right reasons its being a right action is at least to some extent an accident. And if we want our account of the virtuous agent to be such that the ideally virtuous agent *always* performs the overt actions that are morally best, then we need the virtuous agent to always be both aware of and responsive to the right reasons (and unresponsive to the wrong reasons). But in this chapter we are not so much concerned with overt actions as we are with *inner* actions, namely, the virtuous agent's *thinking* about morality. And if we apply the Virtue-Ethical canon to her moral thinking, we get the result that she has the right reasons for maintaining the beliefs she has. The reason the canon properly applies to her inner actions is really no different from the one just given: If we want our account of the virtuous agent to be such that the ideally virtuous agent's thinking and beliefs are

always exactly as they should be, then we need her to always be both aware of and responsive to the right reasons, and unmoved by the wrong reasons.

So we *must* think of the ideally virtuous agent as having perfect justification for all of her moral beliefs. And, to be clear, I mean not merely that her moral beliefs *admit of* perfect justification, but that she *knows* what justifies those beliefs.

What I want to argue next is that an agent who wants to be sure that her moral beliefs are as justified as moral beliefs could be, or at least that she is on her way to that state, will be responsive to ideas about practical laws. Again, these are practical principles that are valid for *all possible agents*; and when I say that a principle is valid for an agent I mean that the agent should conduct herself in accord with it. So, my claim here is that when a human agent is trying to get as much justification as she can about principles' validity for all human agents, she should take seriously questions about principles' validity for all possible agents.

Suppose you believe that some principle, P, is valid for all human agents. But suppose also that you can think of some possible non-human agent for whom, it seems to you, P *is not* valid; rather, it seems to you that some other principle, Q, should be substituted for P in the non-human case. In other words, it seems to you that it would be *morally wrong* for this non-human agent to act/live in accord with P, although human agents' accord with P is *morally required*. Should you bother to try to make sense of this disparity? You might be tempted to think that this would be a waste of time, since, after all, you're a human being, and so only concerned with getting clear on the principles that

are valid for human agents. But unless you can come to understand why P might hold for human agents but not for all possible agents, you cannot claim to fully understand why P holds for human agents. Even supposing that you have some explanation or argument for P's validity for human agents, your inability to make sense of the non-human case shows that your explanation/argument is incomplete. And if your explanation/argument is incomplete, there is some chance that it gets something wrong, and that P in fact *is not* valid for all human agents. An agent who wants to be sure to get things right would be unwilling to take such a chance.

Let me briefly say a few things about resolving such a disparity. Of course, the disparity might be merely apparent. It might be the case that if you fully accounted for the differences between the non-human agent in question and human agents, you would find that P is in fact valid for all of them (and perhaps that Q is not valid for any of them, or instead that Q is in fact entailed by P in the non-human case). But let us suppose you have left nothing out, and that you are right about P and Q. Then there must be an explanation of why, exactly, P is valid for all human agents and an explanation of why Q (but not P) is valid for the non-human agent. The correct and complete explanations will presuppose the same concept of validity and will apply this to the facts about both kinds of agents, respectively, to yield P for the human case and Q for the non-human case. But then each of these explanations is just an application of a complex explanatory principle, call this EP, to a set of particular facts. Applying EP to the facts about human agents yields the result that P is valid for them; applying EP to the facts about the non-human

agent yields the result that Q is valid for her. At this point I can only *assert* that if your explanations are in fact correct and complete, then (a) it is possible to discern a practical law (perhaps more than one) by examining EP, and (b) it can be shown that human agents must act in accord with P in order to act in accord with this (these) practical law(s), whereas the non-human agent must instead act in accord with Q. My hope is that the chapters to come will go some way to support assertions of this sort. I will identify several candidate practical laws and show how some principles, which I claim as valid for all human agents, can be derived from them. Once in a while, I will also note how these practical laws yield different principles for agents that are, in some important ways, quite different from us.

For now, I hope to have made a case for the idea that right reasons for holding some practical principle to be valid for all *human* agents involve considerations about that principle's relations to practical laws, principles that are valid for all *possible* agents. Again, I took up this task in order to argue that the ideally virtuous agent thinks like a Kantian, in the sense that she grounds her beliefs about how she ought to live on her beliefs about practical laws. And, of course, I wanted to get to the related claim that those of us who want as much justification as we can get for our moral beliefs should also be concerned with identifying practical laws. Among these agents are, of course, normal human agents who *strive* for virtue. They should regard the above kind of philosophical activity as a *duty*, more precisely, as part of what it takes to fulfill the duty of *moral self-perfection*, which will be the topic of Chapter Four. That chapter marks the beginning of

my discussion of certain rather general aims that, I will argue, should be adopted by morally committed human agents. Before we get to this, though, I want to explain why the moral commitment, as I have (re)defined it, plays such a central role in my account of virtue and moral goodness.

II. VIRTUE AND THE MORAL COMMITMENT

So far, my arguments in this chapter have had an epistemological flavor; they were driven by questions about how to get more and more justification for one's moral beliefs. The thrust of these arguments could be put as follows: If you want as much justification as you can get for your beliefs about how human agents should act/live, you should make sure to base your beliefs on considerations of practical laws, i.e. practical principles that are valid for *all possible* agents. But on my view, if there were a truly virtuous human agent, she would not think of the practical laws merely as reference points which support her convictions about the principles that are valid for all human agents. Rather, she would see living in accord with the latter principles as nothing more than a means for living in accord with the practical laws.

As the reader may recall, in Chapter One I argued that the moral commitment should be regarded as the *cornerstone* of virtue. My argument there relied on the claim that when the moral commitment prevails in determining an agent's action, conduct in accord with the universally valid practical principles is a necessary consequence. Now, in this chapter I have redefined the moral commitment as the stable, non-instrumental aim of

living in accord with the practical laws—which, again, are the practical principles valid for *all possible* agents. According to the new argument, then, when the moral commitment prevails the necessary result is conduct in accord with the practical principles valid for all possible agents. But now that these details have been made explicit, it might seem less plausible that the moral commitment is necessary for virtue. One could reasonably ask, “Why shouldn’t we remove this ‘cornerstone of virtue’ and replace it with the stable, non-instrumental aim of living in accord with the practical principles that are valid for all *human* agents? After all, when *this* aim prevails, the necessary result is conduct in accord with the practical principles valid for all human agents, and what’s wrong with that?”

An example will help to get my response going. Consider two human agents, Paula and Harry, who fully understand the practical laws as well as what we’ll call the human law: the set of practical principles valid for all human agents (minus the practical laws). They are also equals in terms of strength of will, empirical understanding, and empirical power—exceptionally good with respect to each of these, and also very quick learners. But while Paula is fully committed to living in accord with the practical laws, Harry is fully committed to living in accord with the human law. Now, it might seem that they are equally suited for living in accord with moral requirements. But in fact Paula is better suited than Harry. To see why, notice what would happen if their respective personalities were implanted in two healthy and robust non-human beings, Paula* and Harry*, in an environment where, as it happens, living in complete accord with the human law would

invariably violate one or more of the practical laws.¹⁵ We can imagine Paula* hardly missing a beat, having retained Paula's understanding of and commitment to the practical laws, as well as her exceptional strength of will, and being a quick learner to boot. Harry*, on the other hand, will consistently violate one or more of the practical laws. For, although he has Harry's understanding of the practical laws, he also has Harry's indifference toward them. And let us not forget, the practical laws are, by definition, practical principles that *all* possible agents should accord with, which means they are moral requirements that hold for Harry*.¹⁶

So, my response is as follows. It is, of course, true that when the commitment to the human law prevails, conduct in conformity with the human law is the necessary result. But the above example illustrates that conduct in accord with the human law is not always the same as conduct in accord with moral requirements. Clearly, Harry's suitability for accord with moral requirements is dependent on things external to his

¹⁵ One principle which is plausibly valid for all human agents would be a qualified prohibition on killing, such as the following: Never kill another innocent agent unless (1) she, being fully informed, adequately capable of practical deliberation, and also in a persistent and untreatable state of extreme suffering and/or physical disability, asks to be killed, or (2) this is necessary in order to protect others' interests which outweigh hers. (I leave it to the reader to decide whether (2) is ever satisfied.) But there are non-human scenarios in which an agent who accords with this principle thereby violates the following, which we can plausibly regard as a practical law: Treat all agents with the proper respect and concern. Suppose, for example, that Harry* can greatly improve his fellow non-human agents' health, vitality, and happiness, and radically extend their naturally short lives, by regularly killing and resurrecting them. But suppose also that, in Harry*'s environment, neither (1) nor (2) is ever satisfied.

¹⁶ Introducing such a wild fiction as personality implantation might seem to be, in some sense, cheating. But I could just as well have given an example of some non-human being *born in* the above environment, who gradually and pretty much naturally develops into an agent committed to the human law. (Come to think of it, maybe there is some alien law that Caligula, Hitler, and my former sister in law happened to develop a commitment to—presumably without the aid of alien personality implants!)

personality in ways that Paula's is not. Her personality would do well, morally speaking, in all kinds of scenarios in which a good degree of agency is achievable, and plausibly would never do worse than Harry's if we put them both in any such scenario. And this is all because she has what I have been calling the moral commitment whereas he is committed to the human law. So of course the former commitment is, from a moral perspective, *better* than the latter. In case the reader is not yet quite convinced, let us think again about the ideally virtuous agent. Recall, on my view, she has each of the following in its fullest form: the moral commitment, moral understanding, strength of will, empirical understanding, and empirical power. I have already argued in Chapter One that she never fails to do what is morally best. It should be clear now that this is so no matter what sort of world she is situated in. But this would no longer hold if we were to replace her moral commitment with the fullest commitment to the human law.

It might seem that seeing the moral commitment as *necessary* for virtue has the unappealing result that there are very, very few people who admit of *any degree* of virtue. But what I want to suggest is that there are many human agents who have the moral commitment (to some extent)—precisely as I have defined it here—but are unaware. I suspect that most of the people we could reasonably call “good and decent people” have it. To see this, think of someone whom you consider to be a morally good person, and imagine him/her thrust into a wildly different world where the issues and opportunities for dealing with them are radically different from our own, as in the above example.

We know that people can undergo “changes of heart,” and revise the principles they considered to be *the* fundamental moral principles, just due to (what for them are) novel experiences *here on Earth*. When we encounter situations that seriously challenge our moral convictions, we do not simply give up on morality altogether—at least, not all of us. We often respond by “soul-searching.” That is, we introspect, hoping to find *deeper* commitments which we had all along, albeit tacitly, and which, once brought to light, can help us refine our surface-level way of thinking in a way that accommodates these new challenges. (Notice, this is not unlike the method of resolving disparities described in the previous section.) Of course, people whose life experiences are quite limited are not challenged as often, and so are less often prompted to engage in this kind of soul-searching. We are hardly surprised that these agents claim commitments to principles that, we suspect, they would renounce as too crude, if only they could be brought to think seriously about the right kinds of cases. Nor, then, should we be surprised if very few people who *are* committed to practical principles valid for all possible agents explicitly represent themselves as such. Their life experiences may have been enough to instill the moral commitment without being varied enough or forceful enough to prompt the kind of soul-searching that would bring it to light.

CHAPTER FOUR

ONE'S OWN MORAL PERFECTION

INTRODUCTION

I have now completed my basic account of virtue and moral goodness. In Chapter One, I argued for the idea that virtue consists in an agent's *aptitude for lawful conduct*, which in turn is determined by her possession of what I have called the five moral perfections: the moral commitment, moral understanding, strength of will, empirical understanding, and empirical power. The ideally virtuous agent has achieved each of these perfections in its fullest forms, I argued, and the rest of us are properly appraised by the extent to which we possess them. Chapters Two and Three further developed and defended my account of the moral commitment, the stable, non-instrumental aim of living in accord with practical principles that are valid for all possible agents. In Chapter Two, I argued that we should not think that agents' possession of the moral commitment must be due to free choice in order for them to count as morally good. And although I argued that the ideally virtuous agent has the moral commitment as her *sole* stable, fundamental aim—the *perfect* moral commitment—I also argued that anyone who is to some extent morally committed should be seen as admitting of at least some virtue. On the other hand, in Chapter Three I argued for the view that *only* morally committed agents truly admit of any degree of virtue.

This chapter marks the beginning of my arguments for four, more specific aims that the morally committed human agent will, or at least should, adopt and maintain. Like Kant, I include among these *one's own moral perfection* and *the happiness of others*. I will follow Kant in calling these wide duties, or, duties of virtue, and I will explain what this means shortly. Since I have found no arguments in Kant's texts for the duty of morally perfecting oneself, I will offer my own in this chapter. But I do not think that anything I will have to say here is at odds with Kant's account of this duty. If anything, this chapter adds to Kant's view. However, in later chapters I not only reject Kant's arguments against similar duties concerning *one's own happiness* and *the moral perfection of others*, and provide arguments for such duties, but also argue for an expansion of the Kantian duty concerning the happiness of others (so that non-human animals and other non-agents are included as morally legitimate beneficiaries). All of this is in the service of filling out my account of the more or less virtuous human being.

THE MORAL COMMITMENT AND MORAL IMPROVEMENT

Kant holds that we have a duty to aim at promoting our own moral perfection. In the *Metaphysics of Morals*, he describes this as a part of the more general duty of self-perfection:

This duty [of self-perfection (simpliciter) consists] in *cultivating* one's *faculties* (or natural predispositions), the highest of which is *understanding*, the faculty of concepts and so too of those concepts that have to do with duty. At the same time this duty includes the cultivation of one's *will* (moral cast of mind), so as to satisfy all the requirements of duty. 1) A human being has a duty to raise himself

from the crude state of his nature, from his animality (*quoad actum*), more and more toward humanity, by which alone he is capable of setting himself ends; he has a duty to diminish his ignorance by instruction and to correct his errors. And it is not merely that technical practical reason *counsels* him to do this as a means to his further purposes (of art); morally practical reason *commands* it absolutely and makes this end his duty, so that he may be worthy of the humanity that dwells within him. 2) A human being has a duty to carry the cultivation of his *will* up to the purest virtuous disposition, in which the *law* becomes also the incentive to his actions that conform with duty and he obeys the law from duty. (*MdS*, 6:386-7)

Kant calls self-perfection a *wide* duty to signify that there is no principle that determines how much one is to sacrifice for the sake of perfection, and no set way in which to pursue it.

Several things are missing, however. Kant has not explained why we have the duty of self-perfection—at least not in the *Metaphysics*.¹⁷ He obviously thinks that self-

¹⁷ In the *Groundwork*, Kant presents two arguments for the duty of self-perfection, both of which he seems to have abdicated by the time he wrote the *Metaphysics*. The first asserts that a human rational being cannot consistently will as a universal law (of nature) a maxim to “let his talents rust and be concerned with devoting his life merely to ... enjoyment;” for, “as a rational being, he necessarily wills that all the capacities in him be developed, since they serve him and are given him for all sorts of possible purposes” (*Gl*, 4:423). The second, which invokes the categorical imperative to always treat humanity as an end, is as follows:

[It] is not enough that the action does not conflict with humanity in our person as an end in itself; it must also *harmonize with it*. Now there are in humanity predispositions to greater perfection, which belong to the end of nature with respect to humanity in our subject; to neglect these might admittedly be consistent with the *preservation* of humanity as an end in itself but not with the *furtherance* of this end. (*Gl*, 4:430.)

I find both arguments puzzling. I think the first argument relies on the idea that, for each capacity nature gives us, there is some purpose it is supposed to serve, and that, because of this, we necessarily will that all of our capacities be developed. But the necessary willing claim seems wrong, and it is not clear how it is supposed to follow from the claim that our capacities are given to us by nature for all sorts of purposes. The second argument invokes, seemingly out of nowhere, the idea that treating humanity as an end involves some sort of harmonizing and furtherance of humanity. Kant’s point here, I take it, is that treating rational nature as an end involves not just refraining from undermining humanity (in oneself or in others) but also doing what one can to enhance it. This seems perfectly right. As I will argue in Chapter Eight, however, it is not clear that Kant’s arguments for the basic requirement to always treat rational nature as an end are successful. Perhaps most important for our present discussion, neither argument addresses the question of whether we have a duty to promote our *moral* perfection.

perfection makes one worthy of one's humanity, or, rational nature. But this seems a little flatfooted, and in addition it's not clear whether Kant means to say that this is the reason why we should pursue self-perfection. Is the duty to make oneself worthy of one's own humanity a more foundational duty, on Kant's view, with self-perfection a means of fulfilling it? If so, where is the argument to the effect that we should strive to be worthy of our humanity? And there is a related question. The most obvious reason to think of one's *moral* perfection as worth pursuing is that moral improvement makes one better suited for fulfilling one's moral requirements. And if this exhausts the reasons for pursuing moral perfection, then moral perfection is not to be pursued for its own sake, in which case Kant should say that it is merely an *indirect* duty. (A duty to promote X is an indirect duty, on Kant's view, if X is to be promoted for the sake of something else, Y, which is to be promoted for its own sake. Promoting Y, on the other hand would be a direct duty.) Kant never calls one's own moral perfection an indirect duty, at least not to my knowledge. But if it is an indirect duty, and nothing more, then this is worth knowing. For, in that case, it seems that fulfilling the duty of one's own moral perfection would always have a lower priority than at least one other duty. For example, suppose Kant had it that we are to promote our own moral perfection only because this makes us better at promoting others' happiness, which we are to promote for its own sake. Then it would seem that, on his view, any time one has the opportunity to either promote one's own moral perfection or promote others' happiness (but not both), one should do the latter, all other things being equal.

So, moral improvement's place (or, rank) in Kant's taxonomy of duties is not as transparent as one might like, nor are his reasons for counting it as a duty. But instead of trying to resolve these exegetical difficulties, I would like to present arguments of my own, which I think Kant could, and perhaps in some way does, endorse. My overriding concern, however, is to show not only *that* the morally committed human agent should aim at her own moral perfection, but also *why* she would adopt and maintain this aim.

Let me start with some remarks about the extent to which moral self-improvement is possible. Recall, on my view, an agent's virtue consists in her possession of what I called the five moral perfections: the moral commitment, moral understanding, strength of will, empirical understanding, and empirical power. I have argued that the ideally virtuous agent has each of these perfections in its fullest form, and that it makes sense to morally appraise the rest of us by the extent to which we possess each of them. But I am not convinced that progress with respect *each one* of these perfections is something one can deliberately make happen, and I certainly do not mean to suggest that we should blame people for failing to make it happen. For one thing, I am not completely sure whether it makes sense to say that a person can deliberately strengthen her moral commitment (assuming she has acquired it in the first place—which, by the way, I am not convinced is a matter of choice). Can one deliberately make it the case that over time one has fewer and fewer aims that could conflict with the aim of living in accord with the practical laws? I am *inclined* to agree with Kant (see the above excerpt) that one can, but I certainly have no argument to that effect. I am similarly uncertain on the question of

whether a human being can, in any meaningful sense, deliberately augment her strength of will. But although I want to remain agnostic on the possibility of promoting these two perfections (in oneself), I am much more confident as far as the other three are concerned. It should be uncontroversial that every morally committed human being can deliberately improve her moral understanding, empirical understanding, and empirical power. That is, it is within one's power to improve one's own understanding of the practical laws (and their implications for one's life), one's understanding of one's world, and one's capacities for bringing about the effects one intends to bring about. At least in these senses, improving one's aptitude for lawful conduct is an option for almost all, if not all, morally committed human agents. In saying this, I also mean to reiterate the (Kantian) position that there is always *room* for moral improvement in any morally committed human agent.

So, given that a morally committed human agent *can* morally improve herself, the question now is why she should. As I argued in Chapter One, there is a sense in which it would be irrational for a morally committed human agent *not to* aim at morally improving herself. To be morally committed, as I have defined this, is to have a fundamental (that is, non-instrumental) aim of living lawfully. But any morally committed agent who has not achieved perfect moral understanding, perfect empirical understanding, or perfect empirical power should realize that improvements with respect to these perfections are ways of making herself better suited for lawful living. And this is all she needs to know in order to see the point of further developing those perfections.

For, she then sees that morally improving herself is a necessary means for ensuring greater success with respect to something she values for its own sake.

I should concede, though, that there are certain results one might want established, but which the above argument does not show. For one thing, it does not show that every morally committed human agent should put great effort into morally improving herself. Human agents who have the *perfect* moral commitment will, of course, since they (by my definition) have no fundamental non-moral aims that could conflict with their striving for moral perfection. But recall, in Chapter Two I argued that there is logical space for agents who are morally committed but not perfectly so, and that we should nonetheless call these agents (at least somewhat) morally good. These are agents who *do* have non-moral fundamental aims that can come into conflict with the aim of moral improvement. Some such agents may have a rather weak moral commitment which very often “takes a backseat” to their non-moral aims. So they may hardly ever put effort into morally improving themselves. All the above argument purports to show is that even they will see the point of aiming to improve themselves, and therefore that they will be at least somewhat willing to put forth some effort, however small and however infrequently. Now this might seem a weakness of the view presented so far, since it may seem to demand too little of these not-all-that-morally-good agents. But I think Kant is perfectly correct to say that there is no rational principle which determines how much one should sacrifice for the sake of self-perfection, including *moral* self-perfection (*MdS*, 6:392-3).

A further concession should perhaps be made explicit. My above argument says nothing about agents who altogether lack the moral commitment. Because the argument simply draws out one rational implication of the commitment to living lawfully as found in less-than-ideally virtuous agents, it cannot get off the ground in the case of agents who lack that commitment. But things are not quite as dire as they may seem. Among human agents who do not explicitly aim at *living in accord with practical principles that are valid for all possible agents*, there are of course very many whom we must still describe as aiming to live “a morally good life” for its own sake. They simply do not represent “the morally good life” in terms of accord with practical laws, as Kantians do, and only because of this would not claim themselves as fitting my definition of morally committed agents. But they, too, can see the point of morally improving themselves, for reasons not unlike those of the morally committed human agent. That is, they also can see moral improvement as a necessary means for ensuring greater success with respect to something they value for its own sake. Moreover, their approach to moral improvement would presumably involve promoting their empirical understanding, empirical power, and, hopefully, their moral understanding. (And if my results in the previous chapter are correct, these agents, with sufficient progress in moral understanding, *will* think of “the morally good life” in terms of practical laws—in which case they will count themselves as having the moral commitment.)

Now, there are other kinds of possible human agents that remain unaccounted for so far, namely, those who lack the moral commitment as well as any close cousin of it.

And, frankly, I am not sure that they *can* rationally see a point in morally improving themselves. So I think a very different kind of argument would be needed to settle the question of whether such agents also have a duty to promote their own moral perfection. But these kinds of agents are not my concern, since I am trying to give an account of the more or less virtuous human agent. Some agents who lack anything resembling the moral commitment may conduct themselves lawfully from time to time, and some of them may do this quite reliably. What they all lack, however, is a stable, non-instrumental aim to do what is right, which is central to virtue on any plausible account. Moreover, questions of whether *all* human agents have the duty of moral self-perfection, as Kant claims, are, strictly speaking, orthogonal to my project. My primary goal in this chapter has been just to show that the morally committed human agent has decisive reason to aim at morally improving herself.

CHAPTER FIVE

ONE'S OWN HAPPINESS

INTRODUCTION

Kant holds that each of us has a duty to do what we (tolerably and permissibly) can to promote other rational beings' happiness.¹⁸ But although others' happiness is to be promoted for its own sake, on Kant's view, *one's own* happiness is not. Kant does not say that there is no duty at all to promote one's own happiness, but he calls this an *indirect* duty to signify that the morally legitimate end served by promoting one's own happiness is something other than one's happiness itself. Roughly, Kant holds that being unhappy leads to transgressions of duty, so we are to avoid slipping into this state.

Intuitively, Kant's view seems too inegalitarian: One's own happiness is worth promoting only when one's ability to accord with duty is at risk, but others' happiness is always worth promoting in its own right—not merely when *their* ability to conform with duty is at risk. And so, as long as I am sufficiently confident that I will not violate any duty, I am to promote others' happiness and not my own, on this view, even if they are already very happy and I am not. In this chapter I will critique Kant's reasons for rejecting a direct duty to promote one's own happiness, and offer an argument in favor of

¹⁸ What Kant means by happiness is not straightforwardly discernible from his texts. He sometimes characterizes it in terms of the satisfaction of desires, other times in terms of pleasure/agreeableness and the absence of pain/disagreeableness. And sometimes he seems to presuppose a conception of happiness much like that of the ancient Greeks. But I believe that what I will say here applies no matter which account of happiness one plausibly attributes to Kant.

some such duty. Thus, on my view, the more or less virtuous human agent has, or *should* have, two general *self-regarding* aims, one concerning her own moral perfection, the other concerning her happiness. In later chapters I will argue for two corresponding *other-regarding* duties. Thus, we will arrive at a much more symmetric and egalitarian, but still distinctively Kantian view.

In Section I, below, we will see that Kant offers two arguments against the idea of a (direct) duty concerning one's own happiness. One has it that duties must be concerned with aims that we are reluctant to adopt; and since all human agents have their own happiness as a natural end, it is contradictory to suppose that there is a duty to promote one's own happiness. Kant also argues that there can be no practical law of self-love, that is, one requiring agents to promote their own happiness. This would have to be a "determinate" practical principle which is valid for all possible agents, says Kant. And no such principle is possible, since, for one thing, agents differ in what they need in order to be happy; for another, an agent would have to be omniscient in order to know what *she* needs to be happy. But I will argue that Kant cannot consistently require practical laws to be determinate in the above senses, since he claims at least one practical law which is not. Moreover, there seems to be no good reason to impose this requirement on practical laws in the first place. In Section II, I concede that I am forced to remain agnostic on the existence of a practical law of self-love. But I argue that there may be other practical laws that give rise to a *duty* of self-love—one that satisfies Kant's reluctance requirement.

I. KANT AGAINST SELF-LOVE AS A DIRECT DUTY

In the *Metaphysics of Morals*, Kant argues that, since human beings naturally have their own happiness as an end, and since there can be no duty to do what one already, unreluctantly does, the idea of a duty to promote one's own happiness is incoherent:

For *one's own happiness* is an end that all human beings (by virtue of the impulses of their nature) have, but this end can never without self-contradiction be regarded as a duty. What everyone already wants unavoidably, of his/her own accord, does not belong under the concept of *duty*; for this is a *necessitation* to a reluctantly adopted end. Hence it is self-contradictory to say: one is *obligated* to promote his own happiness with all his powers. (*MdS*, 6:385-6.)

Now, I interpret the above as follows:

- To say that agent A has a duty D to do X is to say that (a) A should do X, AND (b) A is reluctant to do X.
- Human beings are not reluctant to adopt/maintain their own happiness as an end.
- From (1) and (2), human beings have no duty to adopt their own happiness as an end.

Notice that this does not rule out the possibility that there is some practical law of self-love. All Kant has here is an argument to the effect that if there were such a practical law, it still would not give rise to a human *duty* of self-love, since every human agent accords with that law naturally, unavoidably, and unreluctantly. But, although it will not matter much for my discussion below, I should point out that this argument is not obviously precise enough. Supposing all human agents do naturally and unavoidably aim

at their own happiness, it does not follow that they pursue it as much and as well as a law of self-love would require (if there were such a law).

But let us now examine Kant's arguments against such a law. In both the *Groundwork* and the *Critique of Practical Reason*, Kant argues that a requirement to pursue one's own (hedonistic) happiness cannot serve as a practical law, since there is nothing that is conducive to the happiness of *every* rational being (*Gl*, 4:418-9; *KpV*, 5:19-26). For example, he writes,

[S]uppose that finite rational beings were thoroughly agreed with respect to what they had to take as objects of their feelings of pleasure and pain and even with respect to the means they must use to obtain the first and avoid the other; even then they could by no means pass off the *principle of self-love as a practical law*; for, this unanimity itself would still be only contingent. The determining ground would still be only subjectively valid and merely empirical and would not have that necessity which is thought in every law, namely objective necessity from a priori grounds[.] (*KpV*, 5:26)

I think Kant is probably right to say that there is no object that is a source of pleasure for all agents. And so, he would be right to say that no hedonistic principle requiring the pursuit of some particular set of objects holds for all agents. But what about a principle that simply required us to pursue pleasure and to avoid pain (in a way that observes the other moral requirements)? To be sure, this principle of self-love would be indeterminate in the sense that it does not tell us which objects to go after, or how to go after them. And figuring these things out is, quite evidently, by no means straightforward, even with a good bit of experience in one's world and in one's skin. As Kant puts it in the *Groundwork*:

Now, it is impossible for the most insightful and at the same time most powerful but still finite being to frame for himself a determinate concept of what he really wills here [in willing his own happiness]. If he wills riches, how much anxiety, envy and intrigue might he not bring upon himself in this way! If he wills a great deal of cognition and insight, that might become only an eye all the more acute to show him, as all the more dreadful, ills that are now concealed from him and that cannot be avoided, or to burden his desires, which already have given him enough to do, with still more needs. If he wills a long life, who will guarantee him that it would not be a long misery? If he at least wills health, how often has not bodily discomfort kept someone from excesses into which unlimited health would have let him fall, and so forth. In short, he is not capable of any principle by which to determine with complete certainty what would make him truly happy, because for this omniscience would be required. (*Gl*, 4:418)

But Kant cannot consistently regard these problems of indeterminacy as decisive grounds for rejecting a practical law concerning one's own happiness. For, the same problems attach to the requirement to treat rational nature as an end, which Kant claims *is* a practical law (*Gl*, 4:428-9). According to Kant, accord with this law requires, among other things, promoting *others'* happiness and promoting one's own perfection (*Gl*, 4:430). But which of the objects that I can realize will truly promote my neighbor's happiness? Is there anything I can do that is guaranteed not to "backfire"? (Surely if the means to *my own* happiness are chancy in terms of their success, then the means to *others'* happiness are as well.) As for my own perfection, what weaknesses or deficiencies of mine must I deal with in order to truly progress toward virtue, and what sure means for dealing with them are available to me?

In short, complete certainty about what it takes to treat rational nature as an end would also require omniscience. This principle is also indeterminate, evidently more so than the principle of self-love, since the latter only involves promoting the happiness of

one rational being—and, no less, a rational being whom one presumably knows comparatively well. But the indeterminacy of the requirement to treat rational nature as an end does not stop Kant from claiming that requirement as a practical law. Nor, I think, should it. There is nothing wrong with defining practical laws in an abstract way, and then leaving it up to agents to figure out—albeit fallibly—what sorts of things can be done, given the way their world is, to accord with those laws. On the other hand, if Kant wants to maintain that considerations of indeterminacy speak decisively against a law to promote one’s own happiness, he is forced to say that the requirement to treat rational nature as an end is no practical law either.

Kant’s indeterminacy objection against a practical law of self-love evidently is not one he can afford to keep on the books. Moreover, it is not clear why one would think that practical laws, or, universally valid practical principles, must admit of this kind of indeterminacy in the first place. So, for all Kant has said, there could be a practical law of self-love that provides the basis for a human duty to promote one’s own happiness for its own sake. A principle requiring the pursuit of pleasure and the avoidance of pain is just one candidate (one which I find most plausible). Another candidate is one requiring the pursuit of desire-satisfaction and the avoidance of aversion-satisfaction. Contemporary welfarists such as Griffin, Sumner, and Kagan have developed accounts of well-being from which we could generate further candidates. I do not mean to suggest that any of these putative principles of self-love should be regarded as the *sole* practical law. (I

would then be leaning toward some form of ethical egoism.) Rather, I contend that Kant has not shown that the set of practical laws contains no principle of self-love.

II. SELF-LOVE AS A DUTY

I find it difficult to imagine that there is no practical law of self-love at all. This would mean that, from a moral perspective, the importance of one's own happiness is entirely parasitic on the importance of the ends it might serve or compromise. But although I find this hard to believe, I presently have no conclusive argument to the contrary. On the other hand, I am pretty sure that a practical law of self-love cannot be ruled out by any argument either. It seems that a theory must either be agnostic on this question or admit that it is making a stipulation.

Although I take the agnostic option, I need not therefore be agnostic on the existence of a (direct) *duty* concerning one's own happiness. The duty could derive from a practical law that itself is not concerned with happiness. For one thing, there is plausibly a practical law requiring agents to pursue their permissible ends *efficiently*. There is certainly something inherently irrational in doing less than one can in the pursuit of a permissible end when no other end is served by this. And precisely because Kantianism and Virtue Ethics both (rightly) put rationality at the center of morality, this fact provides us with a foothold within either framework.

There can hardly be much doubt that human agents' natural inclinations do not produce the most efficient pursuit of happiness. It is no secret, for example, that we have

a natural tendency to go for immediate gratification even when we reasonably believe that doing so is contrary to our happiness in the long run. Another thing that can make an agent's pursuit of happiness less efficient is her having an inadequate understanding of what tends to make her happy and, as a result, acting blindly on her inclinations more often than she should. In addition, although Kant is right that there are certain delayed effects that we simply cannot foresee, this is not to say that we all already do the best we can in this regard. For these reasons alone, the efficient promotion of one's own happiness requires much more than reliance on one's natural inclinations. But now combine this with the fact that we are all at least somewhat reluctant to do all that is permissible, within our means, and necessary for the efficient pursuit of happiness. The result is that *the efficient pursuit of happiness* fits Kant's definition of a duty. It is a rationally required end which each human agent is reluctant to adopt/maintain.

CHAPTER SIX

THE MORAL PERFECTION OF OTHERS

INTRODUCTION

In chapters Four and Five, I argued for two self-regarding aims that, on my view, the morally committed human agent should adopt and maintain. One of these was the aim of *one's own moral perfection*. In a nutshell, my thesis in Chapter Four was that because a human agent can always improve her aptitude for lawful conduct, it would be irrational for a morally committed human agent *not to* have this as one of her aims. And although I admitted some doubt as to whether it is within a human agent's power to deliberately augment her moral commitment or her strength of will, I maintained that almost all, if not all, human agents can deliberately improve on their moral understanding, empirical understanding, and empirical power. In this chapter, I will be arguing for a corresponding *other-regarding* aim. That is, I will argue that the morally committed human agent should adopt and maintain the aim of helping other agents morally improve themselves.

As I noted at the beginning of Chapter Four, Kant agrees that moral self-perfection is one of the proper aims of the morally good human agent. However, he argues that there can be no duty to have another person's moral perfection as an end, since this would be a requirement to do for others what only they can do for themselves (*MdS*, 6:386).

Although this part of Kant's view has not received much attention, several contemporary Kantians have registered their doubts about it.¹⁹ But it has not yet been shown exactly where Kant's arguments go wrong, nor has anyone to my knowledge shown how Kant could argue for such a duty.²⁰ These are two of my primary goals in this chapter. I will show that Kant's argument presupposes too narrow an account of moral perfection. And

¹⁹ Allen Wood (in Wood, *Kant's Moral Religion* (Ithaca, New York: Cornell University Press, 1970), p.78) writes,

The mutual improvement of men's moral characters through education and religious community play such an important role in Kant's overall view of the moral destiny of man that it is impossible to take as Kant's best thinking the passage cited in which he denies that one can promote the moral good of others. Rather, it seems altogether proper to regard the critical moral philosophy, along with Kant in his most profound moral convictions, as holding that the moral good of *all* finite rational beings is the unqualified and unconditioned end of the finite rational moral agent.

Christine Korsgaard argues that Kant's conclusion is overstated: "Choosing ends on another's behalf is as impossible as it would be disrespectful, but putting others in a good position to choose ends for themselves, and to choose them well, is the proper work of parents, teachers, friends, and politicians" (Korsgaard, "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations," *Philosophical Perspectives* 6 (*Ethics*, 1992), pp.305–332, 331 n. 37). And Lara Denis, whom I believe to have given Kant's texts on this issue the most searching analysis to date, admits a failure "to find any decisive argument to support Kant's claim that others' [moral] perfection is not an obligatory end" (Denis, "Kant on the Perfection of Others" (*Southern Journal of Philosophy*, 1999), pp.21-41, p.21).

²⁰ Wood (1970, p.76) claims that "advice, encouragement, discipline, education, or social improvement" are ways of having "a beneficial effect on the moral character of another, on his 'power to adopt ends in accordance with his own concept of duty'." But he presents no argument, Kantian or otherwise, that bestowing such effects is a duty. He merely argues that Kant lacks good enough reasons for denying the possibility of such a duty. Korsgaard (1992) offers no argument. Denis (1999) argues that Kant should see the duty regarding others' moral perfection as being owed to *all* finite rational beings, in the same way that his duty regarding others' happiness is. As we will see, I do not think this is quite right. In any case, I have some reservations about her argument. She sets it up by noting that Kant's *formula of humanity* (*Gl*, 4:429) requires us to respect ourselves by adopting our own moral perfection as an end, and then asking, "Now if self-respect requires us to recognize our own moral perfection as an obligatory end, why does respect for others not require us to have and to promote the end of others' [moral] perfection?" (p.24). She continues, "the burden of proof falls on Kant: he needs an argument to explain why respect for others does not demand that we include their [moral] perfection among our ends" (p.24). Denis makes a fine case for the claim that "the arguments available to support Kant's claim that the [moral] perfection of others is not an obligatory end fail" (pp.25-35). But she concludes from this that, "Since FH [the formula of humanity] demands that we respect

I will argue that a duty concerning others' moral perfection can be derived from a principle Kant rightly claims as a practical law, the so-called Formula of Universal Law (FUL): Act only on that maxim which you can at the same time will as a universal law (*Gl*, 4:421).

In Section I, I will rehearse and further develop my reasons for thinking that Kant rightly includes moral understanding as a component of moral perfection. And in Section II, I argue that we can promote others' moral perfection by promoting their moral understanding, and so (contrary to Kant) there is no contradiction in supposing that one person can make another person's moral perfection her end. In Section III, I concede that promoting other's moral understanding without violating other duties may require some care, but I argue that this, too, is possible. And Section IV presents several arguments for a corresponding duty. These arguments are available to Kant; they are mere adaptations of his own arguments for the duty to promote others' happiness. As we will see, however, the duty established by these adapted arguments is, in an important sense, less general than Kant's duty regarding the happiness of others. The latter is owed to *all* human agents. All the adapted arguments are capable of showing is a duty to promote the moral understanding of those who want help with their moral understanding. Finally, in Section V, I show that similar arguments can be made for promoting others' empirical knowledge and empirical power.

ourselves by having and promoting the end of our own [moral] perfection, Kant's failure to explain why we do not have a comparable duty with regard to the [moral] perfection of others implies that, in fact, we do have such a duty" (pp.35-6). I'm not sure I agree with Denis' logic here, but in any case even she has not shown that Kant has an argument for the duty in question.

I. KANT ON MORAL PERFECTION AND MORAL UNDERSTANDING

In some texts, Kant comes across as being overly optimistic regarding our ability to determine, in all cases, what the moral duties require us to do. Following his derivation of FUL in the *Groundwork*, he writes:

Thus, then, we have arrived, within the moral cognition of common human reason, at its principle, which it admittedly does not think so abstractly in a universal form, but which it actually has always before it eyes and uses as the norm for its appraisals. Here it would be easy to show how common human reason, with this compass in hand, knows very well how to distinguish in every case that comes up what is good and what is evil, what is in conformity with duty or contrary to duty, if, without in the least teaching it anything new, we only, as did Socrates, make it attentive to its own principle; and that there is, accordingly, no need of science and philosophy to know what one has to do in order to be honest and good, and even wise and virtuous. (*Gl*, 4:403-4)

But even if one were to appeal to FUL as her “compass” for discerning what she morally ought to do, she may nonetheless draw the wrong conclusions or find herself unable to draw any conclusions. What sorts of moral duties, if any, get established by FUL has been a matter of considerable controversy, and it is at least arguable that Kant himself misapplied it. So, even if it is true that FUL is *the* compass for discerning the moral duties, having it in hand obviously does not make discerning them a straightforward matter. And without a correct account of the moral duties, it will be difficult to tackle the further task of figuring out, for any situation one might be in, which of the available actions is/are in conformity with duty. This latter task is one that no human agent can

complete, which is why I maintain that perfect moral understanding is something that human agents can at best approximate.

If Kant was overly optimistic regarding our ability to achieve moral understanding at the time he wrote the *Groundwork*, his attitude had changed by the time he completed *The Metaphysics of Morals*, twelve years later. In his “Introduction to the doctrine of virtue”, Kant argues for two ends that we have a duty to adopt, namely, others’ happiness and our own perfection (*MdS*, 6:386-8). And (as we saw in Chapter Four) under self-perfection, Kant includes not only the strengthening of one’s will but also the development of one’s understanding, which in turn includes one’s moral understanding:

When it is said that it is in itself a duty for a human being to make his end the perfection belonging to a human being as such (properly speaking, to humanity), this perfection must be put in what can result from his *deeds*, not in mere *gifts* for which he must be indebted to nature; for otherwise it would not be a duty. This duty can therefore consist only in *cultivating one’s faculties* (or natural predispositions), the highest of which is *understanding*, the faculty of concepts and so too of those concepts that have to do with duty. ... [The human being] has a duty to diminish his ignorance by instruction and to correct his errors. (*MdS*, 6:386-7.)

Kant’s discussion of virtue in that same work even suggests that he regards the pursuit of moral understanding as something that should never be abandoned:

But virtue is not to be defined and valued merely as an *aptitude* and ... a long-standing *habit* of morally good actions acquired by practice. For unless this aptitude results from considered, firm, and continually purified principles, then, like any other mechanism of technically practical reason, it is neither armed for all situations nor adequately secured against the changes that new temptations could bring about. (*MdS*, 6:383-4.)

Here, Kant is saying that virtue must not be regarded simply as some acquired habit, but as something that results from principles that the agent has thought about and continually refined. If it were possible for a human being to reach a point at which her practical principles were “armed for all situations”, then there would be no need to further refine them; her moral understanding would be adequate. But Kant’s assertion that virtue involves principles that are *continually* purified implies that this work is never completed.

Kant insists that virtue is an ideal that we, as finite rational agents, can *at best* continually approximate (*MdS*, 6:409). And the above seems to suggest that, on his view, this is not just because we never reach a point at which our wills are so strong that there is nothing that can tempt us to act contrary to duty. It is also because we never reach a point at which we perfectly understand (a) the principles that define our moral duties and (b) which of the available actions, for any given situation, accord with duty. In short, our moral understanding can always use improvement. And this is why Kant is right to include a duty to continually cultivate it, as part of the duty of self-perfection.

II. THE POSSIBILITY OF PROMOTING OTHERS’ MORAL PERFECTION

The fact that Kant includes moral understanding as a component of moral perfection is crucial to my critique of his argument against a duty to promote others’ moral perfection. As we will see, his argument is, roughly, that there is a contradiction in having another’s moral perfection as an end, so having this end cannot be morally required. My aim in this section is just to show that, because we can promote others’ moral perfection by

promoting their moral understanding, there is no contradiction in having another's moral perfection as an end. But let me start with a few clarifications, in order to block certain interpretations of Kant's argument. Specifically, I want to rule out certain exegetical claims about what *having another person's moral perfection as an end* involves, to show that certain characterizations of the alleged contradiction are not available to Kant.

First, we should not think that, on Kant's view, my having some other person's moral perfection as an end presupposes that she can *achieve* moral perfection. Rather, he must say that I can have that end even if she will never be morally perfect (and even if I know this). Kant is explicit that moral perfection is an ideal that human beings can at best approximate (*MdS*, 6:409), but this does not stop him from saying that every human being should have *her own* moral perfection as an end (*MdS*, 6:386-7). So, the fact that I cannot make myself morally perfect does not make it contradictory for me to have my own moral perfection as an end, according to Kant. Nor, then, should the fact that I cannot help *another* person become morally perfect make it contradictory for me to have *her* moral perfection as an end. Second, my having another person's moral perfection as an end should not be understood as implying that I can, or believe I can, *directly* make her more morally perfect (e.g., through brain surgery). For, Kant does not require that we be able to directly make others happier, in order for us to have their happiness as an end (*MdS*, 387-8). It is enough that we find ourselves able to do things that *facilitate* their being happier. If he is to be consistent, then, his claim that it is contradictory to have

another's moral perfection as an end cannot be based on the idea that we cannot directly promote others' moral perfection.

So, if I find myself able to help another person make progress toward the ideal of moral perfection, why might Kant maintain that there is, nevertheless, a contradiction in making her moral perfection one of my ends? Here is his argument:

[It] is a contradiction for me to make another's *perfection* my end and consider myself under obligation to promote this. For the *perfection* of another human being, as a person, consists just in this: that he *himself* is able to set his end in accordance with his own concepts of duty; and it is self-contradictory to require that I do (make it my duty to do) something that only the other himself can do. (*MdS*, 6:386.)²¹

According to the above, there is a contradiction in requiring me to do something that I cannot do. But what exactly is it that I cannot do? Kant says that a person's moral perfection *just is* her being able to set her end in accordance with her own concepts of duty. So, perhaps the idea is that (1) I cannot set another person's end in accordance with her own concepts of duty, because, *in general*, end-setting is something no agent can do for another.²² In addition, or alternatively, Kant may be trying to say that (2) an agent's *ability* to set her end in accordance with her concepts of duty is something no other agent can (directly or indirectly) give her or even help her improve.²³

²¹ Given his emphasis on concepts of duty, Kant is clearly concerned with *moral* perfection here.

²² By this point in the *Metaphysics of Morals*, Kant has already argued that it is impossible to make another person adopt an end: "I can indeed be constrained by others to perform *actions* that are directed as means to an end, but I can never be constrained by others *to have an end*: only I myself can *make* something my end" (*MdS*, 6:381).

²³ Denis (1999) has already given a compelling critique of this argument, and much of what I will have to say in this and the next section draws on observations she has already made. Like Korsgaard (1992, p.331, n.37), she is sympathetic to Kant's claim that we cannot set others' ends for them. But, as regards the ability to set one's end in accordance with one's concepts of duty, Denis (1999, p.28) points out that, elsewhere, Kant seems not only to claim that, contrary to (2)

But, even if we assume that (1) and (2) are true, the above argument is importantly incomplete, since its characterization of moral perfection fails to account for the element of moral understanding. As we have seen, Kant elsewhere includes “*cultivating one’s ... understanding, the faculty of concepts and so too of those concepts that have to do with duty*” under the end of self-perfection (*MdS*, 6:387). There, he sees moral understanding as an essential component of moral perfection. In the above, however, Kant asserts that perfection consists *just* in the ability to set one’s end in accordance with one’s own concepts of duty. This implies that (if one has this ability) it does not matter how crude, or how cultivated, one’s concepts of duty are; a person’s ignorance and error regarding the moral requirements have nothing to do with her moral perfection. The problem is not just that Kant is inconsistent; the notion of moral perfection presupposed in the above argument is clearly the inferior one. (Clearly, an agent can be perfect in her ability to set her end in accordance with her own concepts of duty, and yet have concepts of duty that are so crude that she would not count as a morally perfect agent. In my terms, no agent who is seriously deficient with respect to moral understanding can be regarded as ideally virtuous, even if she has achieved each of the other moral perfections in its fullest form.) Moreover, if we keep the better notion in mind, we will see that it *is* possible to promote others’ moral perfection, and that it is possible to have this as an end.

we *can* promote this ability in others, but also to offer an account of how this can be done (*KpV*, 5:159-61; *MdS*, 6:478-84). But, although (2) may be unavailable to Kant, my own critique does not assume it is.

Suppose a student has already adopted the end of living in accord with moral requirements, but is unclear on what they are, and comes to my office hours with questions about some of the moral theories discussed in a course. If our conversation helps her refine her own thinking, then I have helped her improve her moral understanding (regardless of whether that was my intent). Her practical principles may be more refined as a result, thus bringing her closer to the ideal of being “armed for all situations” (*MdS*, 6:383-4). Our ability to exchange ideas makes it the case that we *can* promote each others’ moral perfection, by promoting each others’ moral understanding. And, because this is possible, it is possible to have another person’s moral perfection *as an end*. For, there is nothing contradictory in your aiming to help someone improve her moral understanding. Nor is there anything contradictory in supposing that your reason for having this aim is just that it will help her become more morally perfect.²⁴ If this *is* your reason, you have her moral perfection as an end.

III. THE PERMISSIBILITY OF PROMOTING OTHERS’ MORAL UNDERSTANDING

Let us now briefly consider whether Kant might think it morally impermissible to aim at promoting another person’s moral understanding. What I want to argue here is that it *can be* morally permissible, on Kant’s view, depending on the precise nature of the aim.

²⁴ Here I am assuming that motivational egoism is false.

What matters is whether it is constrained by a proper respect for both the beneficiary and for the other duties one has.

We can get an idea of what the proper respect for the beneficiary amounts to by thinking of the kinds of motives and attitudes that are incompatible with it. In particular, if I am willing to pressure the beneficiary into assenting to my own views about what the moral duties are, I am not respecting her as a *self-legislator*.²⁵ I am willing to forego her making sense, in her own way, of what is morally required of her (and why), for the sake of making sure that she adopts the view I take to be correct. And this means that I am at least to some extent trying to provide her with non-moral incentives to live according to *my* understanding of the moral requirements, rather than appealing to her reason. Also, I am therefore treating her as a means (and so, on Kant's view, it had better be the case that I am also in some way treating her as an end). It seems that, for Kant, a proper respect for the beneficiary entails aiming to facilitate clear, rational reflection without providing her with *any* non-moral incentive for adopting or rejecting any particular view.²⁶ In addition, this aim seems most conducive to promoting moral *understanding*, as opposed to the mere acceptance of a view. And so, unless I have this aim, it is not clear that I am

²⁵ Kant writes, “[T]he will is not merely subject to the [moral] law but subject to it in such a way that it must first be viewed as also giving the law to itself and just because of this as first subject to the law (of which it can regard itself as the author)” (*Gl*, 4:431). And, “the human being is ... subject *only to laws given by himself but still universal*” (*Gl*, 4:432).

²⁶ Kant recommends something like this approach in his discussion of “moral catechisms”, which are conversations in which “the teacher ... methodically draws [answers] from the pupil's reason” (*MdS*, 6:477-84).

really aiming at promoting another's moral understanding, as Kant (rightly) understands it.

As for other duties that can constrain the aim of promoting another's moral understanding, Kant would oppose lying for the sake of getting the beneficiary to accept a given view.²⁷ One also has a duty to oneself, according to Kant, to make sure not to sacrifice so much for the sake of any end that one renders oneself incapable of enjoying "the pleasures of life" (*MdS*, 6:452). So, Kant would say that it is possible to "give too much" in the service of promoting another's moral understanding, and that one has a duty to avoid this. In addition, Kant asserts that a lack of (one's own) happiness can tempt one to transgress one's duties (*MdS*, 6:388). And we can easily imagine how having unmet needs or frustrated desires can lead to impatience or irritability during moral discourse, both of which make it more difficult to treat the beneficiary with the proper respect.²⁸

Promoting another person's moral understanding in a way that is, on Kant's view, morally permissible may require some care. But I see no reason in Kant for thinking that any approach one might take would inevitably violate some duty.²⁹ (Surely Kant didn't

²⁷ Kant has several arguments for an absolute prohibition on lying (see *KpV*, 5:44 and "On a Supposed Right to Lie from Philanthropy", 8:426). But, as I will argue in the next chapter, each of these fails.

²⁸ Of course, we can at least imagine a case in which an agent seeks to promote her own moral understanding just so that she better knows what it takes to live *immorally*! That promoting her moral understanding is impermissible seems obvious enough. To account for this fact, we should further fill out the duty concerning others' moral understanding by saying that it is ultimately concerned with others' moral perfection. And this would mean that there is also no *duty* to promote the moral understanding of those who will never see moral philosophy as anything more than an academic exercise.

²⁹ For more on this, see Denis (1999, pp.31-3).

think that writing philosophical texts about morality is impermissible!) And I submit that there are probably many appropriately respectful approaches.

IV. LAWFULNESS AND OTHERS' MORAL UNDERSTANDING

Let us now set aside questions about the possibility and permissibility of promoting others' moral understanding, and turn to the question of whether Kant has arguments available for a duty. I will start by working toward two arguments, which I adapt from two of Kant's arguments for the duty to promote others' happiness.

In Section II of the *Groundwork*, Kant argues that a maxim *not to promote others' happiness* cannot be willed as a universal law (of nature):

[A]lthough it is possible that a universal law of nature could very well subsist in accordance with such a maxim [of non-beneficence], it is still impossible to **will** that such a principle hold everywhere as a law of nature. For a will that decided this would conflict with itself, since many cases could occur in which one would need the love and sympathy of others and in which, by such a law of nature arisen from his own will, he would rob himself of all hope of the assistance he wishes for himself. (*Gl*, 4:423.)³⁰

According to Kant, in willing the maxim in question as a universal law, one also wills that one not be helped should one ever be in need. The will would therefore “conflict with itself”, says Kant, because one at the same time hopes to be helped should one be in need. (Notice, the argument presupposes that the “hope” for future assistance constitutes a *willing*. For, if it does not, Kant cannot conclude that the will “would conflict with

³⁰ Although the idea of universal laws *of nature* helps Kant make his case (or at least appear to do so) as he argues for other duties in the *Groundwork*, the above argument does not need the “of nature” qualification. So I will henceforth leave this out. Thus, I will be employing the Formula of Universal Law (FUL) itself (Act only on that maxim which you can will as a universal law), and not what might be called the Formula of Universal Law of Nature.

itself’.) So, one cannot consistently will as a universal law the maxim *not to promote others’ happiness*. From this Kant concludes that we have a wide duty to promote the happiness of others, which is equivalent to saying that we have a duty to adopt the end of promoting others’ happiness.

He argues for the same result in *The Metaphysics of Morals*, though in a slightly different way:

The reason that it is a duty to be beneficent is this: since our self-love cannot be separated from our need to be loved (helped in the case of need) by others as well, we therefore make ourselves an end for others; and the only way this maxim can be binding is through its qualification as a universal law, hence through our will to make others our ends as well. The happiness of others is therefore an end that is also a duty. (*MdS*, 6:393.)

For Kant, self-love consists in having the end of one’s own happiness, an end which all finite rational beings have. So, in saying, “our self-love cannot be separated from our need to be loved,” Kant is asserting that we depend on others’ help in our pursuit of happiness. But a maxim that dictates that we receive that help can only be binding, according to Kant, if it is made a universal law that *all* be helped when in need.

Both arguments, then, rely on the idea that, with regard to our own happiness, we are not fully self-sufficient. We sometimes need (and *will*) the help of others, and this is why a general refusal to help others is impermissible, on Kant’s view.

But we can plausibly assume that we are not fully self-sufficient with regard to moral understanding either. As we have seen, agents who have adopted the aim of moral self-perfection, can benefit from other agents at least in the sense that they can improve their

moral understanding through the exchange of ideas about right and wrong. Insofar as they recognize that someone else can help them in this regard, they would want that help. (And recall, in Chapter Four I argued that it would be irrational for a morally committed human agent *not to* aim at morally perfecting herself.) This enables us to construct arguments analogous to Kant's arguments for the duty to promote others' happiness.

Corresponding to the *Groundwork* argument, we have the following argument that an agent who has the end of her own moral understanding cannot will as a universal law a maxim *not to promote the moral understanding of those who will to be helped with their moral understanding*:

It is impossible for an agent who aims at her own moral understanding to will that such a principle hold everywhere as a law. For, in deciding this, her will would conflict with itself, since many cases could occur in which she would need the wisdom of others and in which, by such a law arisen from her own will, she would rob herself of all hope of the assistance she wishes for herself.

In other words, because the agent who aims at moral understanding may at some point need the assistance of others in the pursuit of this aim, she cannot without contradiction will the maxim in question as a universal law.³¹

So, in Kantian fashion, we can conclude that, for agents who aim at moral understanding, there is a wide duty to promote the moral understanding of others who

³¹ Denis (1999, pp.34-35) constructs a similar argument. But her point in doing so is to show that there is no conflict in having one's own moral perfection as an end while also having others' moral perfection as an end. And this, in turn, is part of her rejection of Kantian arguments against a duty regarding others' moral perfection.

also have it as an end. One might object that there are no cases in which an agent who seeks moral understanding *needs* the wisdom of others, at least not in the way that an agent might need the love and sympathy of others in, say, a life-threatening situation. But Kant was not merely concerned with establishing the duty to help those in life-threatening situations. So, this new application of the Formula of Universal Law (FUL) works if we simply construe situations of need as situations in which one wills to be helped. As I noted earlier, this is precisely what Kant needs in the original argument to establish that, in willing non-beneficence as a universal law, the will conflicts with itself. And so the new argument works if, in a similar way, situations could arise for the self-perfecting agent in which she wills to receive the benefit of another's wisdom. But if human agents can never achieve perfect moral understanding, as Kant (rightly) claims, then such situations are always a possibility. Moreover, FUL is a practical law; and, by definition, morally committed human agents aim to live in accord with practical laws. So what we have just seen amounts to an argument for the claim that the morally committed human agent should aim to promote the moral understanding of those who want such aide.

Here is the second one, adapted from Kant's argument in *The Metaphysics of Morals*:

Since the end of one's own moral understanding cannot be separated from the need for instruction through others, human agents who have the end of moral understanding (e.g. *morally committed* human agents) therefore make that end an end for others; and the only way this maxim can be binding is through its

qualification as a universal law, hence through their will to make others' moral understanding their end as well.

Like the argument from which it is adapted, the above presupposes a lack of self-sufficiency, in virtue of which the finite rational being makes one of her ends an end for others, through some maxim. But if this maxim is to be binding, it must be qualified as a universal law, which means that the agent who makes an end of hers an end for others must also make their corresponding end her own. So, if the morally human committed agent's end of moral understanding is to be an end for others she must also, in the case of those who have their moral understanding as an end, make that end of theirs one of her ends.³²

There is an important difference between the duty established by the new arguments and the Kantian duty to promote others' happiness. Happiness, according to Kant, is an end that all finite rational beings naturally have (*Gl*, 4:430), whereas moral understanding is not. Moral understanding and, more generally, moral perfection are ends that we all *ought to adopt*, on Kant's view, but not ends that we all in fact adopt. So, Kant's view is

³² This second new argument, too, appears at first problematic. For, it seems that the end of one's own moral understanding *can* be separated from the need for instruction, since one can do a lot to promote one's moral understanding without relying on others' help. But this problem, if it is a problem, is merely an analog of a problem with the original argument of Kant's, specifically with his claim that "our self-love cannot be separated from our need to be loved (helped in the case of need)" (*MdS*, 6:393). It seems that my self-love, my end of my own happiness, *can* be separated from my need to be helped in the case of need. (We can do a lot to promote our own happiness without relying on what others can do for us.) However, in the case of the original as well as the adapted argument, not much hinges on whether the separation is possible, as long as we assume that the agent makes the relevant end (her happiness or her moral understanding) into an end for others, by some maxim. For, then Kant's assertion that the bindingness of this maxim depends on its qualification as a universal law still leads to the conclusion that the agent herself must make the corresponding end of others her own.

consistent with the idea that there are some agents that *do not* will to have their moral understanding promoted by others or make their moral understanding an end for others. And notice, neither of the two new arguments says anything at all about such agents. Thus, these arguments do not establish that *all* human agents *have* the duty in question. Nor do they establish that the duty is *owed* to all human agents. In effect, they only show that the duty to promote others' moral understanding holds within a subclass of finite rational beings, consisting of those who will to be helped with their moral understanding. Of course, this subclass includes all morally committed human agents who recognize the decisive reasons they have for aiming at their own moral perfection. So, the above arguments do show at least that every morally committed human agent has a duty to help other morally committed agents improve their moral understanding. On the other hand, Kant regards the duty concerning others' happiness as one that is *owed by and to all finite rational beings*.

Now, there *is* an argument available to Kant to the effect that all human agents *have* the above duty. It can be adapted from another one of Kant's arguments for the duty to promote others' happiness. This one comes right after his statement of what is commonly called the Formula of Humanity (FH): "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means" (*Gl*, 4:429).

[C]oncerning meritorious duty to others, the natural end that all human beings have is their own happiness. Now, humanity might indeed subsist if no one contributed to the happiness of others but yet did not intentionally withdraw

anything from it; but there is still only a negative and not a positive agreement with *humanity as an end in itself* unless everyone tries, as far as he can, to further the ends of others. For, the ends of a subject who is an end in itself must as far as possible be also *my* ends, if that representation is to have its *full* effect in me. (*Gl*, 4:430.)

In the above, Kant asserts that the full effect of representing another human being as an end in itself involves making her ends one's own. Kant could also use this idea to argue that *all* finite rational beings have a duty to promote the moral understanding, so far as they can, of those ends in themselves who have chosen their own moral understanding as an end. From the claim that "the ends of a subject who is an end in itself must as far as possible be also *my* ends," it straightforwardly follows that I must (as far as I can) make that subject's moral understanding one of my ends if it is one of hers.

However, for reasons I will make explicit in Chapter Eight, I do not think Kant is entitled to the claim that treating others as ends in themselves entails adopting their ends as one's own. And if Kant were to reject this claim, he may be unable to provide a sound argument to the effect that all human agents have the above duty. I myself have no such argument on hand but, then again, I do not need this result. My general project in this work is to provide an account of the more or less virtuous human agent. The less ambitious result that morally committed human agents have the above duty is still interesting enough. And the two arguments we already have in hand are sufficient for this, given that I have already established (in Chapter Four) that it would be irrational for a morally committed human agent *not to* aim to promote her own moral understanding. She should embrace any permissible means for improving her moral understanding, and

so (according to the above arguments) should be willing to help others who aim to improve theirs.

But should morally committed agents also try to morally enlighten those who *do not* aim at improving their moral understanding? This question is complicated by the fact that agents who think morality is incoherent or too demanding often have no interest in hearing more about it. It is certainly understandable that those who are convinced that there is no sense to be made of what they've heard so far about morality, or that they are bound to fail to live up to its demands anyway, would consider moral discourse a waste of time at best, terribly frustrating at worst. It is also unfortunate, since, at least on my view, morality is neither incoherent nor all that demanding. Certainly if all of our moral duties are *wide* duties, as I will suggest in the next chapter, then all that is *required* is the adoption of certain aims, which one can put as much into as one sees fit. If I am right about this much, morality is not unlike a perfectly reasonable and forgiving person who, in some circles, and through no fault of her own, has acquired a reputation as a harebrained, fetishistic nag. And, keeping with this simile, the question is whether those who know better have some duty to at least clear her name as widely as they can.

I am not quite sure what to say here. On one hand, I am not convinced that all human agents have either the moral commitment or a close cousin of it. So, for all I know, there may be some human agents who have no end or interest that could be served by knowing that morality *does* make sense and *is not* especially demanding. I have no argument for a *duty* to promote *their* moral understanding. In fact, there may be a duty to refrain from

this and put one's efforts into more fruitful endeavors. On the other hand, I am not sure whether we can know other agents well enough to conclude in advance that it would be a waste of time. And if we cannot, then an argument could be made that, in order to "play it safe," one should try. After all, if a *human* agent has the moral commitment, but has been brought to believe that morality is incoherent or too demanding, enlightening her would at least be an indirect way of promoting her *happiness*.³³ And because I think that Kant's original FUL argument above is roughly right, I would say that *any* agent's happiness is something a morally committed agent should aim to promote through permissible means. Still, if we assume that some human agents altogether lack the moral commitment, it seems we would have to know more about their numbers in order to show that we have a duty to *always* "play it safe."

I suspect that the above question is more complicated, perhaps much more, than my remarks suggest. But for now I will be content if I've shown that all morally committed human agents should aim to promote each others' moral understanding.

³³ I defined the moral commitment as a stable, non-instrumental aim of living in accord with practical laws, that is, practical principles that are valid for all possible agents. But, as I argued in chapters One and Three, a human agent can have this commitment without knowing what the practical laws are and/or without knowing what they imply about how she should live. What I am assuming *here* are two claims about things that can make human agents *unhappy*. First, we can plausibly say that having a stable aim that one does not know how to pursue is frustrating—and not merely in a technical sense; for human agents, it can be a source of suffering. Such is the plight of the morally committed agent who believes that there is no sense to be made of morality. She is unable to fill out the content of one of her aims, so of course she does not know how to pursue it. Helping her make sense of things is a way of improving her prospects for happiness, I contend, no matter *how* morally committed she is. Second, feeling that one is not up to the task of meeting standards one genuinely believes one *should* meet is, for most if not all human agents, detrimental to happiness. So, in the case of morally committed agents who have overestimated the *moral* standards, providing the right account could be a huge benefit.

V. LAWFULNESS AND OTHERS' MORAL PERFECTION

As I said in the introduction to this chapter, although my focus has been on promoting others' moral understanding, arguments analogous to the ones I have presented here could be worked out for other components of moral perfection. Now, it is not entirely clear whether we can help others improve with respect to *each* of those components. Recall, in addition to moral understanding, I claimed the following as elements of virtue:

- The moral commitment: The stable, non-instrumental aim of living in accord with practical laws.
- Strength of will: The capacity to resist episodic temptations to neglect the ends one values most.
- Empirical power: The capacity to bring about the things one intends to bring about.
- Empirical understanding: Understanding one's world, including one's own psychology.

I have *some* doubts as to whether one can help another agent improve her moral commitment or strength of will. However, we can obviously help others further develop their empirical power and empirical understanding.³⁴ It should also be obvious that there

³⁴ I do not mean to suggest that one *necessarily* promotes another agent's moral perfection any time one promotes one of the above constituents. It should be clear that if an agent altogether lacks the moral commitment, increases in the other constituents need not morally improve her. Moreover, even if an agent has the *perfect* moral commitment, improving her empirical power or empirical understanding need not constitute an improvement if her moral understanding is seriously deficient. What I do mean to say, however, is that helping an agent improve with respect to one or more of the above traits is *the only way* to help her morally improve herself.

are permissible ways of doing both. (And if there are ways of promoting others' moral commitment and/or strength of will, some of these are probably permissible.)

We can, then, generalize each of the two adapted Kantian arguments to arrive at the more general claim that morally committed human agents should aim to promote each others' moral perfection. I prefer the argument based on the Formula of Universal Law (FUL), since it explicitly employs a practical law. Its more general version would go roughly as follows. Because human agents can never be perfectly suited for lawful conduct, each morally committed human agent would be irrational *not to* aim at morally improving herself. These agents should embrace any permissible means for morally improving themselves, including aide from others. Because of this, none of these agents can will as a universal law a maxim *not* to help others morally improve themselves. So, because morally committed agents should be in accord with FUL (since it is a practical law), each of these agents should aim to promote the moral perfection of other morally committed human agents.

So, this is a duty that all morally committed human agents owe each other. Whether anything like this is owed also to agents who lack the moral commitment is not clear. To be sure, given that I do not think that the moral commitment must be the product of free (supernatural) choice, as Kant appears to, I think it is possible to endow another human agent with the moral commitment. This might be regarded as the first step in promoting the moral perfection of agents who presently are not morally committed. And I do not see how one might argue that this is never permissible. On the other hand, I do not have

an argument to the effect that this is a duty, or even a morally good thing to do.³⁵ Still, as I argued in the previous section, there may still be good reasons to try and improve the moral understanding of those who *evidently* lack the moral commitment, since they may possess it or some close cousin on it after all. In such a case, a little moral philosophy might wind up promoting an agent's happiness and her moral perfection at the same time.

One more thing needs to be made explicit, I think. Given the structure of the above general argument, all that has been established here is a *wide* duty. The above general argument merely shows that it is impermissible for a morally committed agent to adopt a maxim *not to* help other morally committed agents with their moral perfection. All this establishes, then, is that such agents are required to adopt the opposite aim/maxim. In other words, morally committed agents should be willing to make at least *some* effort to morally perfect each other. But *how much* effort is not specified, and I agree with Kant that there is no rational principle that could determine some required degree of effort or sacrifice for these kinds of aims. That is for each agent to decide for herself, for all I have said here.

³⁵ I am developing a view on which there are two kinds of intrinsic value. The first is the value something might have merely as an effect; the other is value had simply as a cause. And my present sense is that the most plausible holder of the first kind of value is pleasure, while the moral commitment itself is the sole holder of the second. However, these arguments are still very much works in progress. Still, if this view could be made plausible, it could provide a starting point for an argument to the effect that instilling the moral commitment in others is, in at least some sense, a morally good thing to do. My primary concern with this line of reasoning so far is that, for all it says, instilling the moral commitment need not be a benefit to the affected agent. In fact, depending on the environment she lives in, it could be detrimental to her prospects for achieving the *other* kind of intrinsic value (and who knows what else). Calling this duty (morally committing others) a wide duty only goes so far; what would be nice is an argument to the effect that, at least as far as agents who lack the moral commitment are concerned, promoting happiness has priority over promoting moral perfection.

VI. TAKING STOCK

Of course, if we accept the arguments from which my arguments in this chapter were adapted, we now have *two* rather general, other-regarding aims that morally committed human agents should adopt and maintain: the moral perfection of other morally committed human agents and the happiness of *all* human agents. And I do think Kant's FUL argument for the latter duty is roughly right, although I will argue in Chapter Eight that he is wrong to think it is owed only to other human agents. Rather, I claim, full accord with FUL entails aiming to promote (through permissible means) the happiness of *any* being that is at all capable of happiness. So, on my view, the morally committed human agent's aim concerning others' happiness should be much more general than Kant's arguments suggest.

So, on my view, there are four general aims that the more or less virtuous human being has in addition to the moral commitment itself: her moral perfection; the moral perfection of other morally committed human agents; the efficient pursuit of her happiness; and the happiness of others.

I believe we have now arrived at a pretty full account of human virtue. That is, I think that if we build the best human being we can, using only the account of the five perfections argued for in Chapter One and the four aims above, we'll have a hard time thinking of anything she lacks *qua virtuous human being*. Let us start by imagining a human being who is as morally committed as a human agent can be. That is, she not only has a stable, non-instrumental aim of living in accord with practical laws, but she has

only one other non-instrumental aim, namely, her own happiness (simpliciter). This she cannot help even if she wanted to, on this Kantian view, since she is still a *human* agent. *But* she pursues her own happiness so efficiently that it never conflicts with her pursuit of her own moral perfection, that of other morally committed agents, and the happiness of others. In fact, we can imagine that the pursuit of these three aims has become *her source* of happiness. Now imagine that she is also as good as a human being can get with respect to moral understanding, strength of will, empirical understanding, and empirical power. Let us call this agent *nomologically* virtuous. She is as close to ideal virtue as a human being can get, on the view proposed here. Is there nevertheless some aim or trait which she might yet lack, but which should be part of any account of human virtue?

My sense is that each of the candidate additions falls into one of two groups. Traits the nomologically virtuous agent would in fact possess *because* she has the above traits; and traits that, though intuitively admirable, really are not defensible as *necessary* constituents of human virtue. Of course, I cannot survey all candidates and the arguments that might be given for them. But I hope that the next chapter (especially its Section III), will give the reader some idea of how I would approach other candidate traits. There, I will argue against Kant's famous claim that lying is *always* morally impermissible. I will critique Kant's arguments and show that in some cases lying is not just permissible, but a *morally good* thing to do. In doing so, I will in effect argue that *honesty* falls into the second group above.

CHAPTER SEVEN

VIRTUE AND HONESTY

INTRODUCTION

Kant presents two distinct arguments for an absolute moral prohibition on lying, understood as intentionally making statements one believes to be untrue. In his famous “On a Supposed Right to Lie from Philanthropy,” he argues that lying undermines the general believability of statements and therefore diminishes the forcefulness of contractually based rights. I call this *the contract argument*. In the *Critique of Practical Reason*, Kant argues that a maxim prescribing the intentional statement of a falsity cannot hold as a universal law of nature. I call this *the universality argument*. I will show, however, that each argument presupposes an extremely exaggerated view concerning the frailty of the general believability of statements; Kant overestimates the adverse effects both of lies and of universalized maxims that prescribe them.

I concede that, given our dependence on each other for information, living in accord with the duties I have argued for in previous chapters most often entails refraining from lying (and other forms of deception). However, I argue there are real human cases in which fulfilling one’s duties *requires* lying. In other words, sometimes truth-telling is flat out wrong, and lying is *lawful conduct*. So, I contend, the morally committed human agent who has made adequate progress in morally perfecting herself, particularly with

respect to moral and empirical understanding, *will not* have an absolutist's attitude against dishonesty. Instead she thinks of honesty merely as a mode of conduct that *tends* to be lawful. I conclude that honesty is not a necessary constituent of human virtue.

I. THE CONTRACT ARGUMENT

In his famous "On a Supposed Right to Lie from Philanthropy", Kant argues that even when one cannot evade answering a question, one still is not morally permitted to be untruthful:

Truthfulness in statements that one cannot avoid is a human being's duty to everyone, however great the disadvantage to him or to another that may result from it; and although I indeed do no wrong to him that unjustly compels me to make the statement if I falsify it, I nevertheless do wrong in the most essential part of duty *in general* by such falsification, which can therefore be called a lie (though not in a jurist's sense); that is, I bring it about, as far as I can, that statements (declarations) in general are not believed, and so too that all rights which are based on contracts come to nothing and lose their force; and this is a wrong inflicted upon humanity generally.³⁶

What makes lies impermissible, according to this argument, is the adverse effect they have on the *believability* of statements in general, and hence on contractually based rights. The idea is that if contracts are to be binding, the statements on which they are based must be believed. But if it is not the case that statements generally are believed, contracts will not bind, and so the rights based on contracts will have no force.

Now, surely Kant does not mean to suggest that telling any kind of lie is sufficient to *completely undermine* the believability of statements and the forcefulness of contractually

³⁶ Kant, "On a Supposed Right to Lie from Philanthropy", 8:426; see Gregor (1996), p.612.

based rights. For, obviously, we have at least some tendency to believe others' statements and to expect others to honor their contracts, even though we are all aware that people sometimes lie. Perhaps Kant is trying to say that, in lying, one *lowers* the extent to which statements are believed and, as a result, the extent to which contracts are expected to be honored. But this is not always the case. If, for example, one makes sure that the lie one is about to tell will not be exposed as a lie, the general believability of statements, etc., may go completely unscathed. So, Kant would still need to explain what is wrong with lying in cases where there is no threat to the general believability of statements, etc.

But there is another important sense in which the contract argument is incomplete. It presupposes that there are no duties which could, in some situations, *require* one to tell a lie *despite* the risk of diminishing the general trust in statements and contracts. In effect, Kant is saying that, from a moral perspective, there is nothing more important than maintaining the general trust. (Let us not forget, Kant explicitly denies the permissibility of lying for the sake of preventing murder.) But this seems terribly wrong-headed, so Kant certainly would need to explain why preserving/promoting the general trust has the priority he alleges.

To be sure, doing what one can to preserve and promote the general believability of statements and the forcefulness of contractually based rights seems to be an important moral aim. Our abilities to communicate, to propose contracts, and to bind ourselves through them significantly improve our potential for learning, cooperation, and mutual

advantage. We can safely suppose that maintaining a general trust in statements and contracts serves many moral ends. The question Kantians should be asking, however, is whether the general trust is *an end in itself*, that is, something that is worth preserving and promoting for its own sake (and not merely a means for achieving moral ends). For, if it is not, undermining it can be permissible and, in some cases, required. But it seems Kant is in no position to claim the general trust as an end in itself, since he is explicit that there is only one end in itself: *rational nature* (*Gl*, 4:428). And if the general trust is not an end in itself, treating rational beings as ends in themselves would appear to have priority over preserving and/or promoting the general trust. In fact, Kant is explicit that all of our duties *are derivable from* the requirement to treat rational nature as an end in itself (*Gl*, 4:429). So, it seems he would have to admit that in those rare cases where lying is the only available way to treat rational nature as an end in itself, this is what one ought to do. In fact, as I will explain in Section III, cases can—and do!—arise where one must lie to an agent in order to treat *her* rational nature as an end in itself.³⁷

To summarize, the contract argument seems to exaggerate both the frailty and the importance of both the general believability of statements and the forcefulness of

³⁷ This fact is also problematic for attempts to defend Kant's position by invoking his idea of a *Kingdom of Ends*, an ideal state of mutual respect between all rational beings (*Gl*, 4:433-7). Kant is explicit that we are to act only on principles that could serve as laws in a Kingdom of Ends (*Gl*, 4:436). And one might suppose that, at least for Kant, there can be no deception in a Kingdom of Ends, and therefore that it is wrong to act on any principle that prescribes deception. But there *can be* deception in a Kingdom of Ends, if it is possible to deceive a rational being for the sake of treating her as an end in itself. For, Kant is explicit that the laws of a Kingdom of Ends are grounded entirely on the fundamental requirement to treat rational nature as an end in itself (*Gl*, 4:433). On his view, it is *precisely* when all rational beings respect each other as ends in themselves that a Kingdom of Ends is achieved.

contractually based rights. For, it is possible to lie while preserving the general trust. And there appear to be duties which Kant himself recognizes and which could, in some cases, require an agent to lie despite the damage this would do to the general trust.

Furthermore, it is not as if the contract argument succeeds when applied to other forms of deception. For, even if we set aside questions about the importance of the general trust, the fact remains that we often find ourselves in a position to deceive others while preserving the general trust. And, for all the contract argument says, any kind of deception that preserves the general trust could be morally permissible. But, supposing there *are* moral ends that are more important than preserving/promoting the general trust, there may be cases in which one is morally required to deceive another even if one knows this would guarantee a decline in the general trust.

II. THE UNIVERSALITY ARGUMENT

On the one hand, he argues that, since a maxim prescribing the intentional statement of a falsity cannot hold as a universal law of nature, such statements are not morally permissible:

When the maxim on which I intend to give testimony is tested by practical reason, I always consider what it would be if it were to hold as a universal law of nature. It is obvious that in this way everyone would be necessitated to truthfulness. For it cannot hold with the universality of a law of nature that statements should be allowed as proof and yet be intentionally untrue. (*KpR*, 5:44)

The key premise in the above, expressed by the last sentence, is that there could be no universal law of nature that lies “be allowed as proof”. This is supposed to explain why acting on a corresponding maxim is morally impermissible, and thus why the only permissible kind of testimony is truthful testimony. Unfortunately, Kant has omitted a few crucial details. It is unclear what the precise content of the maxim in question is, and it is therefore also unclear why Kant thinks it cannot hold as a universal law. So we need to do our best to fill these gaps in order to evaluate the argument.

Kant’s specification of a maxim always includes an action or a deliberate omission. And it is clear enough from the above that the action in this case is *making an untrue statement*, telling a lie. Usually, however, he also includes a description of the agent’s situation and/or the aim she intends her action/omission to serve. The universality argument omits a description of the situation, presumably because Kant is after the general result that lying is *in all situations* morally impermissible. And we can only guess that he takes the aim of the lie to be: getting someone else to accept one’s statement “as proof”. It is not clear how strong a notion of “proof” Kant presupposes here. He would obviously be wrong to say that, in lying, we generally intend our claims to stand as *conclusive* evidence for the truth of a false proposition. In order for an agent to see a point in lying, it is sufficient that she expects or hopes that her false statement will incline her audience to give her the benefit of the doubt. More plausibly, then, Kant means to say that, in lying, one intends one’s statement to be taken as evidence (simpliciter) for the

truth of a false proposition. The maxim, then, must be: I will lie for the sake of evidencing to others a false proposition.

Now, why Kant thinks this maxim “cannot hold with the universality of a law of nature” is far from obvious. The only explanation that readily suggests itself is that Kant here presupposes a line of reasoning he has employed elsewhere, in an argument against the moral permissibility of making a false promise for the sake of acquiring a loan one has no intention of repaying.³⁸ The line, suitably adjusted, would have it that a law of nature that people tell lies whenever they aim to evidence false propositions could not sustain itself. For, if people always lied when they wanted to evidence false propositions, it would become common knowledge that people do this, which in turn would only ensure that people would no longer believe any statements at all. And then one could no longer evidence a false proposition by telling a lie, since no one would take statements as evidence anymore. Thus, telling a lie would no longer be a way of evidencing a false

³⁸ In *Groundwork of the Metaphysics of Morals* (4:422), Kant argues:

Another finds himself urged by need to borrow money. He well knows that he will not be able to repay it but sees also that nothing will be lent him unless he promises firmly to repay it within a determinate time. He would like to make such a promise, but he still has enough conscience to ask himself: is it not forbidden and contrary to duty to help oneself out of need in such a way? Supposing that he still decided to do so, his maxim of action would go as follows: when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen. Now this principle of self-love or personal advantage is perhaps quite consistent with my whole future welfare, but the question now is whether it is right. I therefore turn the demand of self-love into a universal law and put the question as follows: how would it be if my maxim became a universal law? I then see at once that it could never hold as a universal law of nature and be consistent with itself, but must necessarily contradict itself. For, the universality of a law that everyone, when he believes himself to be in need, could promise whatever he pleases with the intention of not keeping it would make the promise and the end one might have in it itself impossible, since no one would believe what was promised him but would laugh at all such expressions as vain pretenses.

proposition. In effect, this law of nature would eliminate itself, since it would make such deception impossible.

Assuming this is how Kant intended to be understood, the universality argument, like the contract argument, appeals to the general believability of statements. Unlike the contract argument, however, it *does not* rely on the idea that maintaining the believability of statements is morally important in its own right. Rather, Kant's point here is that, because the above lying maxim would, as a universal law, completely undermine the believability of statements, lying would then no longer be a way of evidencing false propositions. Thus, the lying maxim "cannot hold" as a universal law. According to the universality argument, then, the impermissibility of lying follows from the impermissibility of acting on maxims that cannot hold as universal laws. So, obviously, one cannot object to the universality argument on the grounds that it exaggerates the importance of maintaining the general believability of statements. For, it makes no claims about this at all, nor does it rely on any such claim. It does, however, rely on a claim about the *frailty* of the general believability of statements. For, it presupposes that no statements would be believed if the above lying maxim were to become a universal law of nature. And this, as I will show, is precisely where the argument goes wrong.

The maxim in question prescribes lying for the sake of evidencing false propositions. The corresponding universal law of nature, then, would make it the case that, whenever a person has decided to evidence a false proposition, she will lie. To be sure, if this became a law of nature, and if everyone knew it, the general believability of statements

would probably suffer. But *just how much* it would suffer is not clear. Notice, even if everyone knew about this law of nature, this does not mean that they would also know when a lie was being told. And so, even if the frequency of lies increased significantly as a result of this law of nature, the frequency of *exposed* lies might only increase a little. But it is the frequency of *exposed* lies that affects the general believability of statements, not the frequency of lies. So, there is no reason to think that, under this law of nature, people would no longer believe anything they were told. Nor, then, is there any reason to think that it would be impossible (with such a law in place) to evidence a false proposition by lying. In other words, it seems that, contrary to the universality argument, the above lying maxim *can* hold as a universal law of nature.

The problem with the universality argument, if I've interpreted it correctly, is that it (like the contract argument) exaggerates the frailty of the general believability of statements. It incorrectly assumes that the lying maxim would, as a universal law of nature, have the result that people would no longer believe anything they were told. Similarly, in order to apply the universality argument to some *other* form of dishonesty, one would have to assert that the relevant (dishonest) maxim would, as a universal law of nature, completely undermine some kind of believability or trust. (For, if I have things right, the key idea in the universality argument is that when the relevant kind of trust is completely undermined by the new law of nature, the relevant form of dishonesty can no longer serve its purpose.) But it is not clear that any such assertion is correct.

III. LAWFUL VERBAL CONDUCT

I submit, therefore, that both the contract argument and the universality argument are unsound. Because the latter appeals to the idea of universal validity, it is the kind of argument that should draw a morally committed human agent's attention. (Again, I have defined these agents as having a stable, non-instrumental aim of living in accord with *practical laws*, i.e. practical principles valid for all possible agents.) So, I want to say a few more things about it, in order to arrive at a view of how morally human committed agents *should* think about lying, truth-telling, and verbal conduct in general.

As I noted in the previous section, the universality argument is not very specific about the content of the lying maxim. It does not mention anything about the agent's situation, nor is it very explicit about the aim that the act of lying is intended to serve—except perhaps that one's statement “be allowed as proof.” I suggested that Kant's reason for omitting these details might have been that he was after the very general result that lying is, in *all* cases, morally impermissible. *If* his goal was to show that lying is morally impermissible *in all possible cases*, I hope I have already shown that his argument falls short. In other words, Kant has not shown that there is a practical law against lying. But even so, there's a sense in which I may still have been too charitable in my interpretation of the universality argument, by allowing the aim served by lying to be specified simply as *evidencing a false proposition*. To be sure, I think this is always part of the motive behind (intentional) lies, but this is rarely the whole story. In the overwhelming majority of cases, the agent has some further aim in mind which she expects to be *served by*

evidencing a false proposition. And it would be a huge mistake to think that, in every case—even in every *human* case—the ulterior motive *must* be a bad one.

As we saw in Section I, Kant's contract argument failed to show that it is wrong to lie in murderer-at-the-door cases. I seriously doubt that any argument can succeed. A more Kantian than myself might respond that, even granting that such a lie would be the lesser of two evils, it is an evil all the same, since it involves treating an agent (the murderer) merely as a means. But even this overstates things, since I can lie to you while at the same time doing something else that constitutes treating you as an end in itself. Further argument would be needed to show that lying *necessarily* involves treating one's audience *merely* as a means no matter what else one does at the same time. And, although I won't go into this here, I am pretty sure that Kant's theory cannot accommodate such an argument. More importantly, though, there are cases in which lying to someone is quite obviously the *best* way of treating *her* as an end in itself.

For example, suppose a doctor is quite sure that her patient is in such bad shape emotionally and physiologically that full disclosure of material diagnostic and prognostic information would literally kill him. Suppose also that the physician has gotten consent from a surrogate decision-maker for a treatment that would almost completely restore the patient's health. But just as she is wheeling him into the surgery room, he begins to get agitated, and asks, "Doc, shoot me straight, you've told me everything, right?" Knowing that he is bound to arrest and most likely die if she doesn't settle him down at once, what kind of action is most consistent with treating him as an end in itself? My answer: She

not only needs to lie, she really needs to “sell it.” (When he’s in recovery and stable enough, she can explain it all.) That, on my view, is not just permissible in this case; it is morally *good*. And it is what a morally committed physician with adequate moral and empirical understanding would do. Although I would cash this out primarily in terms of the physician’s best option for promoting her patient’s happiness, I think any Kantian would have to admit that this the physician’s best option for treating her patient (and even her patient’s rational nature itself) as an end in itself.

Human cases in which lying is a morally good thing to do are probably quite rare. That this is so is evident if we think just about what it takes to promote other human agents’ happiness. Because we can rarely if at all do this directly, one of the best things we can do is to better enable others to pursue their own happiness. Making sure that others know enough, and that they are not misguided about matters that are likely to affect them, is very often *essential* for this. (To fully appreciate this, one only needs to imagine how far any of us would get if human beings suddenly were *not at all* able to communicate!) And it probably goes without saying that lying is rarely a good way to promote one’s own or another’s moral perfection.

But these are just facts about how often lying is and is not conducive to fulfilling one’s duties. As such, I contend, they do not constitute reasons for thinking that morally committed agents should place any *special* emphasis on honesty. On my view, a human agent who is as close to virtue as a human can get thinks of honesty in much the same way she thinks about non-violence, politeness, and civil obedience—in much the same way I think about driving on the

right-hand side of the road. Although these modes of conduct are very often appropriate, this is only because they very often are the best available options for acting in accord with practical laws. Should she find herself in a situation where some form of dishonesty is the best option (as in the above case), I contend that she would not hesitate. Moreover, I think it is a mistake to suppose that she should feel any sense of compunction when she is dishonest. For, those who properly understand when lying is and is not lawful, and whose moral commitment and strength of will are sufficient to ensure that they take no liberties, have no need of the common sentiments about honesty and dishonesty; these sentiments could only impair them. Of course, morally committed human agents who have not yet reached that point should perhaps be careful to make sure that, when they lie, their motives are the right ones and their assessments of their situation are adequate. But they should be careful about all sorts of things, and in the same ways. I see nothing in Kant or elsewhere that convincingly shows anything especially problematic about lying. Because of this, I do not count honesty among the traits that the morally committed human agent should have.

CHAPTER EIGHT

HUMAN VIRTUE AND CONCERN FOR NON-AGENTS

INTRODUCTION

In Chapter Six, I argued that the morally committed human agent should adopt and maintain two other-regarding aims: other human agents' happiness and other morally committed human agents' moral perfection. In both cases, I relied on one of Kant's arguments based on the Formula of Universal Law (FUL): Act only on that maxim that you can at the same time will as a universal law (*Gl*, 4:421-3). In this chapter, my primary goal is to argue that the aim concerning others' happiness should be broadened so that *non-agents* are also counted among its putative beneficiaries. Here, too, I will be employing FUL. But although I am thereby siding with Kant in thinking of FUL as a practical law (a practical principle valid for all possible agents), I am at the same time rejecting a good number of Kant's explicit positions—some of which are understandably regarded as essential to any truly Kantian account of virtue. For one thing, Kant holds that all of our duties concerning the treatment of non-rational beings are *indirect* duties; they are entirely parasitic on duties owed to rational beings (*MdS*, 6:443). In other words, treating non-agents one way or another is morally worthwhile only when it benefits some agent(s) in a certain way. This is precisely what I want to deny here. But Kant's position on this is just a consequence of a more foundational claim. He claims that *all* of our

duties are derivable from the Formula of Humanity (FH), which requires us to always treat “rational nature” as an end, and never merely as a means (*Gl*, 4:439). As far as I can tell, there is no sound way of deriving any *direct* duty toward non-rational beings from FH. So, if FH in fact *has* the status Kant awards it, there simply *cannot be* any direct duties toward non-agents. Moreover, Kant claims that FH and FUL are really one and the same practical law (*Gl*, 4:436). And if this is right, then I would be wrong to think that a direct duty owed to non-agents can be derived from FUL in the first place. So before I present my positive argument, I have a good bit of path-clearing to do. Most of this chapter is devoted to critiquing Kant’s thought about the Formula of Humanity.

I will start by showing, in Section I, that the direct duty concerning other agents’ happiness cannot be derived from FH. The key point there will be that, given Kant’s account of “rational nature,” treating rational nature as an end *does not entail* promoting its possessors’ happiness. Kant has given several other arguments for this duty, as we saw in Chapter Six, but he undercuts them all by insisting that each of our duties can be derived from FH. All told, preserving FH’s status renders Kantian morality almost entirely indifferent to the happiness, well-being, and suffering of *all* beings, rational or not. As we will see in Section II, however, Kant’s case for FH’s all-duty-grounding status leaves a lot to be desired. Kant comes to his position by arguing for the following two theses: (a) only an end in itself can ground a practical law; (b) only rational nature is an end in itself (*Gl*, 4:428-9). We will see that his arguments for (a) and (b) are importantly incomplete. Moreover, I will argue that the conjunction of (a) and (b)

contradicts Kant's claim that the Formula of Universal Law (FUL) is a practical law. The key point of Section II, then, is that Kant is not entitled to (a) or (b), and he *needs* to retract at least one of these. But once this is done, he can no longer claim that an alleged duty must be derivable from the Formula of Humanity in order to be a legitimate duty. Kantians should consider it sufficient if a duty can be derived from FUL. And in Section III, I attempt such a derivation of a (direct) duty to aim at promoting the happiness (or, perhaps more precisely, the well-being) of *non-rational* beings.

The positive argument aims to more precisely specify the direct duty concerning others' happiness I claimed in Chapter Six. I maintain that Kant's FUL derivation is successful. But I point out that, especially given my observations in Section II, Kant has not shown that duties derived from FUL are owed only to agents. I go further, though, and use FUL to show that the morally committed agent should not discriminate against non-agents as putative beneficiaries—except, of course, when it comes to things she cannot possibly do for a non-agent. At the very least, I will argue, FUL requires her to aim at promoting the *well-being*, hedonistically construed, of both rational and non-rational beings. (And on my view, as well as Kant's, well-being is at least a constituent of happiness.)

I. HUMANITY AND HAPPINESS

If my arguments in this and the next section are correct, Kant's view of our moral requirements concerning the treatment of non-human animals is just one symptom of an

important problem at the foundational level of Kantian Ethics as we know it. My goal in this section is to provide a fuller account of this symptom and others. Again, I aim to show that if we maintain that all duties are derivable from the Formula of Humanity, as Kant does, then we wind up with a morality that is virtually indifferent to the happiness, well-being, and suffering not only of non-agents, but of agents as well. Let us start with some observations about Kant's take on morality and non-rational beings.

Kant does argue that all human rational beings have a duty to refrain from "violent and cruel treatment of animals", but only because he assumes that such treatment "dulls [one's] shared feeling of their suffering and so weakens and gradually uproots a natural predisposition that is very serviceable to morality in one's relations with other people" (*MdS*, 6:443). For Kant, the moral constraints on our treatment of animals rest entirely on a claim about what Robert Nozick calls "moral spillover": Given our psychology, treating animals poorly makes us less sympathetic *in general*, and so makes us worse at treating people well.³⁹ But, as Nozick and several others have pointed out, this claim is not obviously true of all of us.⁴⁰ At the very least, one could plausibly say that *some* human agents can compartmentalize whatever malicious attitudes or dispositions they have toward animals, so that these do not interfere with their treatment of their fellow human beings. It is not clear how Kant could plausibly maintain that they should be kind to animals. Furthermore, if we searched long enough we might even find someone who

³⁹ Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), 35.

⁴⁰ Ibid. Also see J. Skidmore, "Duties to Animals: The Failure of Kant's Moral Theory", *The Journal of Value Inquiry* 35 (2001): 541-59; Peter Carruthers, *The Animals Issue* (New York: Cambridge University Press, 1992), 160.

is vulnerable to aggressive impulses which impair her ability to sympathize with her fellow human beings, but which she could easily vent by torturing animals every now and then. Kant is committed to saying that she *should* do this.⁴¹ Strictly speaking, he is committed to the view that whether one should be kind or cruel to animals depends only on what it takes to improve one's aptitude for treating other rational beings as one should.⁴²

It might seem that, assuming the overwhelming majority of us *are* susceptible to “moral spillover” in the way Kant thinks, the practical implications of his view are not so different from those of, say, theories that award moral rights to animals.⁴³ But the differences become clearer once we account for the fact that one—perhaps *the*—major cause of harms visited on animals is our buying behavior. Very often, when we pay for animal products we thereby fund practices which visit great harms on animals. Claims about moral spillover do not say much against these consumer actions. To be sure, it is at least arguable that a human agent who allows herself to “turn a blind eye” toward her buying behavior's indirect consequences for animals runs the risk of becoming blind-eyed in a more general way. Even so, psychological compartmentalization is much easier here than it is when one directly and deliberately harms an animal. Moreover, I suspect

⁴¹ This example is adapted from one by Allen Wood. See Wood, “Kant on Duties Regarding Nonrational Nature,” *Proceedings of the Aristotelian Society* Supplement 72 (1998): 189-210, esp. 194-5.

⁴² Kant seems unable to avoid a similar view concerning the treatment of *human* non-agents. See Wood (1998), 198; O'Neill, “Kant on Duties Regarding Nonrational Nature II,” *Proceedings of the Aristotelian Society* Supplement 72 (1998): 221.

⁴³ See Dan Egonsson, “Kant's Vegetarianism,” *Journal of Value Inquiry* 31 (1997): 473-83; O'Neill (1998); Lara Denis, “Kant's Conception of Duties Regarding Animals: Reconstruction and Reconsideration,” *History of Philosophy Quarterly* 17 (2000): 405-23.

that, presently, those who profit from animals tend to make it their business to limit public awareness of their practices. In effect, they reduce the threat of “moral spillover” without diminishing the harms we visit on members of other species. Animal rights advocates would hardly call this progress, but it seems that card-carrying Kantians *must*.

And let us not forget, it is the “shared feeling of [others’] *suffering*” that Kant claims as the “natural predisposition that is very serviceable to morality in one’s relations with other people” (*MdS*, 6:443, my italics). We can see why it might be important to avoid weakening or uprooting this predisposition, if morality in one’s relations with other people is defined in no small part by a direct duty to promote other agents’ *happiness*.⁴⁴ But, as I will show in a moment, no such duty can be derived from the Formula of Humanity (FH). So, if Kant is right to say that all duties are derivable from FH, there is no direct duty concerning other agents’ happiness. And this would significantly attenuate the reasons for preserving or cultivating one’s capacities for sharing in others’ suffering. The Kantian reasons for being kind rather than cruel to non-human animals would, as a result, lose much of their force.

So, let us now turn to Kant’s attempt to derive the duty concerning other agents’ happiness from FH. Following his statement of FH in the *Groundwork*, Kant argues that

⁴⁴ What Kant means by “happiness” is a matter of some controversy. He sometimes describes it in terms of the satisfaction of inclinations, i.e. habitual sensuous desires (*Gl*, 4:399, 405; *KpV*, 5:73; *MdS*, 6:212); other times in terms of agreeableness (*KpV*, 5:22, 60). He also defines *well-being* in terms of agreeableness and freedom from disagreeableness (*ibid*), which seems to imply that happiness and well-being are one and the same, for Kant. At first glance, it is unclear whether all of these characterizations can be reconciled. But I will ignore this for now, and simply assume that Kant’s use of the term denotes either inclination-satisfaction or agreeableness.

the “*full effect*” of representing a human being as an end in itself involves adopting her ends as one’s own, including her happiness:

[C]oncerning meritorious duty to others, the natural end that all human beings have is their own happiness. Now, humanity might indeed subsist if no one contributed to the happiness of others but yet did not intentionally withdraw anything from it; but there is still only a negative and not a positive agreement with *humanity as an end in itself* unless everyone tries, as far as he can, to further the ends of others. For, the ends of a subject who is an end in itself must as far as possible be also *my ends*, if that representation is to have its *full effect* in me. (*Gl*, 4:430.)

What makes human agents ends in themselves, on Kant’s view, is their rational nature (or, humanity), that is, their abilities to reason, to act on principles, and to freely set ends for themselves. The key point in the above argument, then, is that the “full effect” of representing another human being as one who possesses these rational capacities includes adopting her ends as one’s own (to the extent that one can). Once we combine this with the idea that *all* human agents have the end of their own happiness, we get the result that the full effect of representing any human agent as an end in itself entails adopting her happiness as an end of one’s own.

Let me explain why I think this argument fails. For one thing, I do not think Kant has shown that we have the more general duty to adopt other human rational beings’ ends as our own. We might be tempted to believe that respecting a human being’s rational capacities involves adopting the ends she has set for herself through the exercise of those very capacities. But even this requires an argument, for it is not unreasonable to ask why the “positive agreement with humanity as an end in itself” involves anything more than

protecting and enhancing others' rational capacities. In general, the fact that an agent possesses some capacity is not by itself a good enough reason to do anything with or for the products of her exercise of that capacity. So, even if we were to grant that rational nature is the sole fundamental moral value, we could still reasonably ask why any of its products are, from a moral perspective, worth promoting. And I have found no compelling answer in Kant or in the secondary literature.⁴⁵

It is especially difficult to see how Kant could, without contradiction, use the Formula of Humanity to make a case for adopting others' *happiness* as an end. For, he claims happiness as a *natural* end for all human beings, and not as one we set for ourselves

⁴⁵ According to Guyer (2000, 150-71), Kant holds that "our capacity to set and pursue ends of our own choice as a fundamental manifestation of our freedom," and our moral duties are all derivable from the basic requirement to do "what is necessary to preserve and enhance the existence and exercise of freedom itself." But Guyer admits that Kant has no successful argument (in any of his texts) for the idea that freedom is the sole fundamental moral value; at bottom, Kant simply asserts this as an indemonstrable but undeniable fact (162-71). There are further problems with the view Guyer attributes to Kant (though I do not mean to imply that there is a problem with Guyer's exegesis). First, it is not clear how the *pursuit* of one's freely set ends is necessarily a "manifestation of freedom." Once I have freely set an end for myself, what prevents my pursuit of that end from becoming habit rather than a continued exercise of my freedom? Second, even if we agree that treating other rational beings as ends in themselves also includes enhancing *their exercise* of their freedom, it remains unclear how this, in turn, entails adopting the ends they have set as one's own. I can do what I can to make sure that a fellow rational being is well-equipped to pursue whatever ends she has or will freely set, and at the same time refuse to pursue those ends myself.

Other scholars read Kant as attempting to ground this duty (to adopt others' freely set ends as one's own) on the idea that we must, if we are to be consistent, attribute value to others' ends, since we must regard the capacity to set ends (no matter in whom we find it) as *value-conferring*. (See Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), 106-132; Wood, *Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999), 130.) On this view, I necessarily attribute value to my freely set ends, but this is only possible if I also attribute value to my end-setting capacity, in which case I must also attribute the same value to others' end-setting capacities, and so to *their* freely set ends. As Adrienne Martin points out, however, there are good reasons to doubt the validity of the inference from our valuing of the ends we set to our valuing of the end-setting capacity (Martin, "How to Argue for the Value of Humanity," *Pacific Philosophical Quarterly* 87 (2006): 96-125, esp. 110-3).

through the exercise of our rational capacities. It is not entirely clear that Kant can consistently claim happiness as an *end* in the first place.⁴⁶ Supposing he can, it remains unclear how the “full effect” of representing another human rational being as an end in itself entails adopting her happiness as an end. For, it is unclear how her possession of this end figures into a representation of her *qua end in itself*, that is, as a being having the capacities that constitute rational nature. The conceptual connections between these capacities and the natural end of happiness seem tenuous at best. On the one hand, we have the capacity for reason, the capacity to act on principles, and the capacity to set ends in a way that is independent of natural causes. On the other, we have an end consisting in either the satisfaction of habitual sensuous desires or the achievement of agreeableness (or, pleasure). Certainly my representations of each of the rational capacities and the end of happiness become connected in a minimal sense when I represent a human being as possessing them all. But my representation of her as an end in itself need not include her end of her own happiness, just as it need not include her eye-color or shoe-size.

To be sure, protecting/promoting another’s rational capacities may often require promoting her happiness. But, again, the duty Kant is trying to establish here is a *direct* duty; it involves aiming to promote others’ happiness *for its own sake*—and not merely for the sake of preserving or enhancing their rational capacities. And the above argument

⁴⁶ Robert Johnson argues that squaring Kant’s claim that happiness is a natural end with his claim that “To have any end of action whatsoever is an act of *freedom* on the part of the acting subject, not an effect of *nature*” (*Metaphysics of Morals*, 6:385), is difficult if not impossible. See Johnson, “Happiness as a Natural End”, in M. Timmons, ed., *Kant’s Metaphysics of Morals: Interpretive Essays* (New York: Oxford University Press, 2002), 317-30.

does not show that this duty is derivable from the requirement to treat rational nature as an end in itself. Moreover, there is no satisfactory argument that *could* show this, if rational nature just is the cluster of the aforementioned rational capacities and happiness is defined in terms of inclination-satisfaction and/or agreeableness. For, one can treat another human being's rational capacities with the utmost, active respect and concern without caring at all about her sensuous inclinations or her prospects for agreeableness for their own sake.

It might be objected that I have misunderstood Kant's account of rational nature, and that (even if Kant's own derivation fails) the duty to promote other human rational beings' happiness *can* be derived from the requirement to treat rational nature, properly understood, as an end in itself. For example, one might contend that the possession of sensuous inclinations itself is a necessary element of rational nature, on Kant's view.⁴⁷ Thus, one might continue, treating another's rational nature *in its entirety* as an end in itself involves doing what one can to satisfy her sensuous inclinations (as well as aiming to protect/promote her rational capacities). And so, if we construe happiness in terms of the satisfaction of sensuous inclinations, as Kant often seems to, then treating another's rational nature as an end in itself *does* entail trying to contribute to her happiness for its own sake after all.

The problem with this proposal, though, is that Kant is unable to include the possession of sensuous inclinations as an essential component of rational nature. For, he

⁴⁷ Wood (1998, p.203) interprets Kant in this way.

claims that *all* rational beings are ends in themselves (*Gl*, 4:428, 436; *KpV*, 5:87). And some of the beings he counts as rational beings altogether lack sensuous inclinations. Here I am thinking of *holy wills*, (non-human) rational beings which Kant describes as incapable of acting in violation of moral requirements—*precisely because* they lack sensuous inclinations (*Gl*, 4:412-14; *KpV*, 5:32; *MdS*, 6:383, 405).⁴⁸ Kant must, if he is to be consistent, count holy wills as ends in themselves, since he counts them as rational beings. But then, since Kant insists that rational nature is the *only* end in itself, he must grant that holy wills possess it, in which case he *cannot* claim the possession of sensuous inclinations as an essential part of rational nature.

This brings us to a general exegetical point that I believe will be worth keeping in mind. Given that Kant thinks both that all rational beings are ends in themselves and that rational nature is the only end in itself, he is committed to the view that all rational beings possess rational nature. Thus, any alleged element of Kantian rational nature that *is not* shared by all Kantian rational beings (finite rational beings, non-finite rational beings, good wills, holy wills, and even *evil* rational beings) would land Kant with contradictions. So, any interpretation that attributes to Kant an account of rational nature that involves more than the capacities for reason, principled action, and free end-setting is problematic. This forces us to regard human beings' rational nature as something that is very much disconnected from their natural end of happiness, given Kant's account of each. And it is precisely because of this, I contend, that the direct duty to aim at

⁴⁸ Kant even makes room for *finite* holy wills—that is, finite rational beings that altogether lack sensuous inclinations (*MdS*, 6:383).

promoting other human rational beings' happiness for its own sake cannot be derived from the requirement to treat rational nature as an end in itself. In other words, this duty is not derivable from the Formula of Humanity.

Of course, as we saw in Chapter Six, some of Kant's arguments for the duty in question make no (explicit) appeals to the requirement to treat rational nature as an end in itself. If at least one of these arguments is successful, the problem I have identified here may seem benign. After all, one might say, Kant's taxonomy of duties remains intact. But recall, Kant holds that *all* duties can be derived from FH. And if my above critique is correct, this claim is at odds with any argument alleging a direct duty concerning other agents' happiness. So, Kant must either give up on this duty or deny that all of our duties are derivable from the Formula of Humanity.

Taking the first option would have serious implications for Kantian morality. We know Kant is committed to the counterintuitive view that whether one has a duty to be kind to non-rational beings or, instead, a duty to be cruel to them is entirely dependent on what it takes to make oneself better at treating people well. Given the all-duty-grounding status Kant awards FH, he is committed to a similarly counterintuitive view regarding our treatment of other *rational* beings: Whether one should promote another person's happiness or instead diminish it depends entirely on which is more conducive to the promotion of her (or someone else's) rational capacities. On this view, the only time I am morally required to aim at promoting another agent's happiness is when this is *the best way* to protect or improve her (or someone else's) ability to reason, ability to act on

principles, and/or ability to freely set ends. As long as there are other equally effective means for promoting rational nature, I need not be concerned about others' happiness at all. And, of course, this view also has it that one should make other agents *unhappy*, even cause them to suffer, in cases where this is most conducive to protecting or enhancing rational capacities. So, whether one is morally required to promote some other agent's happiness or instead required to promote her suffering is, on this view, a question that must be answered on a case by case basis.

And this result, in turn, weakens Kant's "moral spillover" argument against cruelty toward animals. Recall, he relies on the idea that a disposition to share in others' suffering "is very serviceable to morality in one's relations with other people" (*MdS*, 6:443). This would be somewhat of an exaggeration if the moral treatment of other human agents did not entail aiming to prevent or diminish their suffering, and sometimes required promoting it. A disposition to share in others' suffering would not always be an advantage, and often would make it more difficult to focus on treating rational nature as an end in itself. So, even if we supposed that cruelty toward animals *necessarily* diminishes this disposition in human agents, we still could not base a case against such cruelty on the Formula of Humanity.

As I see things, if all duties are derivable from the Formula of Humanity, as Kant claims, the only thing we have a duty to promote for its own sake is rational nature itself. Aiming to promote or prevent anything else is morally required when and only when this is most conducive to protecting or enhancing one's own or other agents' rational

capacities. Such is the fate of happiness and suffering, no matter whose, if the Formula of Humanity has the status Kant awards it. And it is unclear whether rational beings' freely set ends would fare any better.

II. THE STATUS OF THE FORMULA OF HUMANITY

My goal in this section is to critique Kant's arguments for the all-duty-grounding status of the Formula of Humanity (FH). I certainly think Kant is right to say that any alleged duty must be derivable from *some* practical law in order to count as a legitimate duty. But I aim to show that Kant's case for the idea that *all duties are derivable from the Formula of Humanity* is importantly incomplete and problematic. As I mentioned earlier, his case depends on the following two theses: (a) only an end in itself can ground a practical law; (b) only rational nature is an end in itself (*Gl*, 4:428-9). I will start by showing that Kant's argument for (a) is incomplete. I will then argue that the conjunction of (a) and (b) is incompatible with Kant's claim that the Formula of Universal Law is a practical law. Then I will show that Kant's argument for (b) is also incomplete. Given all of this, I contend both that Kant is not entitled to (a) or (b), and that he *must* retract at least one of these. And, as I will explain, this means that there is no good reason for thinking that each of the duties that can be derived from the Formula of Universal Law *must be owed only to rational beings*.

We begin, then, by examining Kant's argument for the claim that (a) only an end in itself can ground a practical law:

The ends that a rational being proposes at his discretion as *effects* of his actions ... are all only relative; for only their mere relation to a specially constituted faculty of desire on the part of the subject gives them their worth, which can therefore furnish no universal principles, no principles valid and necessary for all rational beings and also for every volition, that is, no practical laws. Hence all these relative ends are only the ground of hypothetical imperatives.

But suppose there were something the *existence of which in itself* has an absolute worth, something which as *an end in itself* could be a ground of determinate laws; then in it, and in it alone, would lie the ground of a possible categorical imperative, that is, of a practical law. (*Gl*, 4:428.)

As the “and in it alone” qualification in the final sentence makes clear, Kant takes himself to have shown that if there is such a thing as a practical law, then it must be grounded on some end in itself. When the worth an agent attributes to an end is due to (sensuous) desires of hers that are not common to all rational beings, this end is merely a relative end, which cannot ground a principle that is valid for *all* rational beings. In order for a practical principle to be valid for all rational beings, it must be grounded on something whose very existence has “an absolute worth,” an end in itself. Kant goes on to assert that (b) only rational nature is an end in itself.

Let me point out first that Kant has nowhere ruled out the possibility of an end which is neither an end in itself nor one that gets its worth from its possessor’s idiosyncratic desires. In fact, he seems to admit that there is at least one such end: that of *living in accord with universally valid practical principles* (*Gl*, 4:401-2). Moreover, I am not sure that Kant can count *happiness* as a “relative end,” given the way he characterizes relative ends in the above. If happiness is a natural end, it does not fit the description of an end that a rational being “proposes at his discretion” (although the things we pursue *for the*

sake of happiness do). Nor is it obvious that the worth we attribute to happiness itself is due to our desires. Granted, we have seen that Kant sometimes construes happiness in terms of the satisfaction of habitual desires (*Gl*, 4:399, 405; *KpV*, 5:73). But even if the satisfaction of desires is what constitutes happiness, it does not follow that a rational being values happiness as an end only because she possesses desires. (Why couldn't a rational being that lacks desires, e.g. a holy will, attribute worth to *others'* happiness?) Finally, let us not forget, Kant also defines happiness in terms of agreeableness (*KpV*, 5:22, 60). And whether agreeableness satisfies the criteria for being a relative end is also unclear.

So, in order to maintain that (a) only an end in itself can ground a practical law, Kant would need to show that none of the above ends can ground a practical law. To be sure, he has argued that *one's own happiness* cannot ground a practical law, his reason being that it is impossible for a human rational being to produce "any principle by which to determine with complete certainty what would make him truly happy, because for this omniscience would be required" (*Gl*, 4:417-9. See also *KpV*, 5:21-6). And, obviously, a similar argument could be made in the case of *others'* happiness. But Kant's argument presupposes that *any end* that is to be the ground of a practical law must be such that it is possible for a human rational being to "determine with complete certainty" what it takes to truly promote it. And this criterion precludes *rational nature* from being a ground of a practical law. For, no human being can be completely certain about what it takes to promote her own or others' rational capacities, much less what it takes to do all the other

things Kant claims are involved in treating rational nature as an end in itself—which *include* promoting happiness! So, Kant cannot consistently use this criterion to rule out happiness as a ground of a practical law.

At the very least, Kant's argument for the claim that (a) only an end in itself can ground a practical law, is importantly incomplete. But it is worth noting that, even before he presents that argument, he has argued for the practical law known as the Formula of Universal Law: "act only in accordance with that maxim through which you can at the same will that it become a universal law" (*Gl*, 4:397-402, 421). And that argument makes no appeal to the existence of any end in itself. So, at least at first glance, it seems that Kant admits at least one practical law which *does not* depend on the existence of any end in itself. Now, if Kant were to count *maxims' conformity with universal law* as an end in itself, he could maintain that Formula of Universal Law is grounded on an end in itself after all. But Kant claims that (b) only rational nature is an end in itself. And once we combine (a) and (b), we get the claim that rational nature is what grounds all practical laws, including the Formula of Universal Law.

So, in order to maintain the Formula of Universal Law as a practical law, Kant needs to show that the mere existence of the capacities that constitute rational nature (in oneself and/or in others) does in fact ground the requirement to act only on maxims one can will as universal laws. To be sure, it certainly seems that possessing these capacities (for reason, principled action, and free end-setting) makes one *capable* of choosing one's maxims for action by applying a criterion of universal lawfulness. And one could

reasonably suppose that a being that lacks them *is not* capable of making such choices. But if these capacities are to be seen as the ground of the Formula of Universal Law, it is not enough to say that possession of them is precisely what makes obedience to that imperative possible. Kant must explain why the rational capacities themselves are precisely what make that obedience *obligatory*. And the question is not just how those capacities all by themselves generate an obligation, but how they manage to generate the specific obligation *to act only on maxims one can will as universal laws*. Now, we know Kant thinks that rational nature is an end in itself and that we have an obligation to treat it as such. Perhaps he thinks treating it as such *entails* acting only on maxims that one can will as universal laws. But, as we saw in Section I, Kant has not explained how it is that treating rational nature as an end in itself goes beyond aiming to protect and/or enhance the rational capacities. And unless it *does* go beyond this, the above entailment claim presupposes this entailment claim: aiming to protect and/or enhance one's own and others' rational capacities entails acting only on maxims one can will as universal laws. But, for one thing, there is no successful argument for this latter entailment claim in Kant.⁴⁹ For another, it seems that it cannot be right. Acting only on universalizable

⁴⁹ As Guyer (2000, 153-9) points out, there are some texts in which Kant seems to suggest that compliance with the Formula of Universal Law is necessary for the maximal realization of the absolute moral value of freedom in oneself and others. Guyer writes:

On this approach there is no initial assumption of the moral value of adherence to the moral law for its own sake; rather, freedom per se is intrinsically valuable, but the introduction of consistency both within one's own choices of ends and among the ends of oneself and others by means of compliance with the requirement of universalizability is argued to be necessary in order to maximize the exercise of this freedom and thus maximally realize its potential intrinsic value. On this account, freedom itself is the absolute value and adherence to the law is the condition necessary for the maximal realization of this value rather than a part of its very concept. (156)

maxims involves much more than aiming to protect and/or enhance one's own and others' rational capacities.

So, Kant cannot claim that rational nature is what grounds the Formula of Universal Law. And this means that he must rescind either the thesis that (a) only an end in itself can ground a practical law, or the thesis (b) only rational nature is an end in itself. Unless he gives up on one of these, he is forced to deny that the Formula of Universal law is a practical law.

We have already seen that Kant's argument for (a) is importantly incomplete, so he could easily retract it. I will now show that the same goes for (b). I will divide (b) into two claims: (b1) rational nature is an end in itself; (b2) there are no further ends in themselves. I aim to show that Kant has no compelling case for (b1) or (b2).

As we have seen, Kant claims that an end in itself is something whose existence has "an absolute worth," and not merely a worth that rests entirely on the possession of

If this is Kant's view, he certainly does hold that aiming to protect and promote one's own and others' rational capacities—specifically, their free end-setting capacity—entails acting only on maxims one can will as universal laws. Adherence to the Formula of Universal Law is required *only because* that adherence maximally realizes the sole fundamental moral value, which inheres in our wills' independence from the laws of nature (155). But problems with this view are not restricted to worries about the fact that (as Guyer well knows) Kant only gets to it by stipulating the absolute value of freedom as an indemonstrable but undeniable fact. One further concern is that, given that freedom is understood in terms of independence from natural causes, it is unclear how any person's actions could enhance another's freedom. (How can any of my overt actions enhance your freedom, if the effects of these actions are just natural phenomena?) So it is unclear how freedom's value generates all of the positive other-regarding duties Kant claims we have. Second, according to Guyer, this view rests on the assumptions that freedom unconstrained by any law is a terrible thing (158), and that the only non-natural law is Kant's moral law (155). The former assumption contradicts the idea that freedom is the sole fundamental moral value. The latter assumption is suspect, given Kant's failure to show that there can be no law grounded on one's own happiness. Such a law would be a non-natural alternative to Kant's moral law.

idiosyncratic desires, which is the case with “relative ends” (*Gl*, 4:428). We also saw that, for Kant, only an end in itself can ground a practical law, a practical principle that is valid for all rational beings. At least at first glance, it appears that all Kant is saying is this: In order for something to be a ground of a practical law, it must be the case that *every rational being values her possession of it*. So, perhaps having “absolute worth” *just is* being valued by every rational being, for Kant, and perhaps this is the sole requirement for being an end in itself. But notice what this requirement leaves out. It does not say that something must be valued *for its own sake* by every rational being in order for it to be an end in itself. And it might be thought that any account (of what it is to be an end in itself) that omits this qualification is far too crude to be Kant’s own. After all, if something is to be an end in itself, it seems quite obvious that it must first be an *end*. And when Kant says that something is an end for a rational being, it seems only natural to read him as saying that this rational being values that thing for its own sake. In addition, leaving out the for-its-own-sake qualification would land Kant with the view that something could be an end in itself even if some—and even if *all*—rational beings value it only as a *means*.

However, requiring that an end in itself must be something that every rational being values for its own sake spells trouble for the thesis that (b1) rational nature is an end in itself. For one thing, Kant rightly admits that there are possible rational beings, e.g. holy wills, which are in the most literal sense *completely* dedicated to living in accord with the practical laws. It seems that the only thing these rational beings could value for its own

sake is acting in accord with the universally valid practical principles. To be sure, such beings would have good reason to place great value on their rational nature, since this is something without which they could not even ascertain the precise content of those principles, let alone figure out what it takes to act in accord with them in particular situations. As noted earlier, the possession of rational nature is precisely what makes deliberate conduct in accord with practical laws possible. So, there is an important sense in which one would understate things by saying that a holy will thinks of her rational nature as a *mere* means. For, her rational nature is *indispensable* to the pursuit of any and all of her ends. Nonetheless, it seems that Kant would contradict himself in claiming that holy wills value their rational nature *for its own sake*.⁵⁰ Moreover, it seems Kant would be wrong to say that all *human* rational beings value their own rational nature for its own sake. It does seem plausible that all human rational beings value—or, at least, have good reason to value—their rational capacities in much the same way that holy wills do, namely, as indispensable means for the pursuit of their ends. But I think that many (perhaps most) human rational beings, if asked, would say that this exhausts the value of rational nature. It might be suggested that these rational beings are simply failing to recognize the decisive reasons for valuing rational nature as more than an all-purpose, indispensable means. But Kant has not shown that there are decisive reasons.

⁵⁰ Kant's various claims about *autonomy* and *self-legislation* might be taken as compelling evidence that, on his view, the value a rational being places on acting in accord with the moral law is entirely parasitic on the absolute value she attributes to her rational nature. But I think Guyer (2000, 162-71) is right to conclude that Kant has no argument showing that each rational being necessarily attributes absolute value to her rational nature. On Guyer's reading, Kant ultimately asserts that rational nature's absolute value is undeniable.

So, it would be implausible and, evidently, inconsistent for Kant to say that every rational being values her own rational nature for its own sake. And he has not made a compelling case for the idea that each rational being *should* value her own rational nature for its own sake (and should not value it merely for its usefulness as a means). Because of this it is entirely unclear why one should think that rational nature, not just one's own but any other being's, is valuable *in itself*.

But then Kant cannot be regarded as successful in his attempt to show that (b2) there are no further ends in themselves. We can readily agree that the list of candidates Kant considers is far from exhaustive:

All objects of the inclinations have only a conditional worth; for, if there were not inclinations and the needs based on them, their object would be without worth. But the inclinations themselves, as sources of needs, are so far from having an absolute worth, so as to make one wish to have them, that it must instead be the universal wish of every rational being to be altogether free from them. Thus the worth of any object *to be acquired* by our action is always conditional. Beings the existence of which rests not on our will but on nature, if they are beings without reason, still have only a relative worth, as means... (*Gl*, 4:428).

And I should point out that Kant merely *asserts* that non-rational beings “have only a relative worth, as means.” More importantly, though, Kant evidently has not specified all of his criteria for something's being an end in itself. All we know is that things that have merely conditional/relative value cannot be ends in themselves, nor of course can things that every rational being *disvalues*. But these (negative) criteria do not rule out all other candidate ends in themselves, e.g. pleasure/agreeableness, desire-satisfaction, life, and

virtue. A case for (b2) would have to tell us what it is that distinguishes rational nature from all other candidates. Kant clearly has not done this.

In short, Kant's case for the thesis that (b) only rational nature is an end in itself leaves too much unexplained. We have seen that his argument for the thesis that (a) only an end in itself can ground a practical law is also inadequate. In addition, we saw that rational nature cannot ground the Formula of Universal Law, which means that Kant cannot consistently maintain both (a) and (b) while claiming FUL as a practical law.

Again, I think Kant is right to say that any alleged duty must be derivable from *some* practical law in order to count as a legitimate duty. My aim here was to show that Kant's case for the idea that *all duties are derivable from the Formula of Humanity* is importantly incomplete and problematic, due to its reliance on both (a) and (b). Nothing I have said here contradicts the idea that some duties are derivable from the Formula of Humanity. But, given that rational nature grounds the Formula of Humanity but not the Formula of Universal Law, there may be some duties that are derivable from the Formula of Universal Law but *not* derivable from the Formula of Humanity. To be sure, Kant claims that these two practical laws are really one and the same, but the fact that their respective grounds differ already gives us good reason to doubt this. Moreover, since the Formula of Universal Law is not grounded on the idea of rational nature as the sole end in itself, there is no reason to think that each of the duties that can be derived from the Formula of Universal Law *must be owed only to rational beings*. This will come up in

the next section, where I derive a duty to promote others' well-being (for its own sake) from the Formula of Universal Law.

III. WELL-BEING AND THE FORMULA OF UNIVERSAL LAW

The Formula of Universal Law (FUL) states, "Act only in accordance with that maxim through which you can at the same time will that it become a universal law" (*Gl*, 4:421). As we saw in Chapter Six, the direct duty concerning other agents' happiness can be derived from FUL, roughly as follows. As human agents, we necessarily aim at our own happiness and will for others to promote it should we ever need their assistance. Because of this, none of us can consistently will as a universal law a maxim *not to* promote other agents' happiness. (We would be willing both that no one promote anyone else's happiness and that others promote our happiness.) According to FUL, then, we should not act in accordance with this maxim. Rather, each of us should be willing to make at least *some* sacrifices for the sake of promoting other agents' happiness. Thus, FUL establishes a duty to aim at promoting other agents' happiness—not for the sake of promoting their rational nature or anything else, but simply for its own sake (which is what makes this a *direct* duty).

In the previous section, I argued that there is no reason to think that each duty derived from FUL must be owed only to agents. Of course, when it comes to duties requiring us to promote things that only agents are capable of, it would be contradictory to claim that they are also owed to non-agents. And so, depending on how we construe *happiness*, it

might be absurd to say that there is any duty—direct or indirect—concerning non-agents’ happiness. I think Kant can consistently say that non-agents (non-rational beings) are capable of happiness. But this is a fairly controversial exegetical issue which does not need to be settled here. Instead, I will show that accord with FUL requires us to aim at promoting the *well-being* of both agents and non-agents, where well-being is defined in purely hedonistic terms, following Kant (*KpV*, 5:22, 60).

I take it to be uncontroversial that human agents aim at their own well-being, understood as pleasure and the absence of pain. To be sure, it seems quite obvious that many human agents have other aims which they regard as much more important, and which play a much greater role in determining their actions. But the aim of achieving pleasure and avoiding pain is one that no human agent can completely abdicate. Moreover, none of us aims at well-being *just because* well-being is good for something else. We try to achieve pleasure even when we expect no further advantage from it, and we certainly avoid pain just for the sake of avoiding pain. We aim at well-being *for its own sake*. So we can see how an argument almost identical to the above FUL derivation could be constructed for a direct duty to aim at promoting other agents’ well-being.

The question here, though, is whether a derivation could take a more general form, so that the duty established is owed to *all* beings that are capable of well-being, or, all *sentient* beings. What I want to do is start with a rough argument, modeled on the above FUL derivation, which I intend as a *prima facie* case for such a duty. I’ll then address a rather obvious shortcoming of that argument and show how it can be dealt with through

further applications of FUL.

The rough argument is as follows. As human agents, we necessarily aim at our own well-being and hope that others will promote it—for its own sake—should we ever need their assistance. Because of this none of us can consistently will as a universal law a maxim (M1) not to promote other *sentient beings*' well-being for its own sake. (Henceforth, I will omit the “for its own sake” clause.) For, we would then will both that no agent promote the well-being of another sentient being, and that other agents promote our well-being. According to FUL, then, we should not act in accordance with M1. Rather, each of us should be willing to make at least some sacrifices for the sake of promoting other sentient beings' well-being.

The problem is that there appears to be at least one alternative to M1 that each of us *can* consistently will as a universal law. Consider M2: I will not promote the well-being of sentient *non-agents*. Since none of us is a non-agent, it seems each of us can consistently will that no one promotes the well-being of non-agents. So, from the fact that we cannot consistently will M1 as a universal law, it *does not* follow that we should be willing to promote sentient non-agents' well-being. If the above prima facie case is to stand, it needs the help of an argument against the likes of M2. It needs to be shown that a human agent who refuses to aim at promoting non-agents' well-being violates FUL even if she aims to promote agents' well-being.

Of course, we should first consider human agents who have adopted a maxim like M2 because they believe that only the well-being of agents is *morally* worth promoting.

Presumably some of them believe, as Kant evidently did, that aiming at promoting agents' well-being is constitutive of respecting their rational nature, but that nothing like this is owed to non-agents. Now, we have already seen problems with this way of thinking in Section I. The connections between rational nature and pleasure and pain are not robust enough to support the idea that respecting rational nature entails aiming to promote its possessors' well-being. But the question here is whether a maxim that presupposes some robust connection(s) of this sort can nonetheless be willed as a universal law. So, let us ignore Section I's results for a moment, and consider the maxim (M2a): I will promote others' well-being only out of respect for their rational nature. In willing M2a as a universal law, I would be willing for other agents to promote my well-being only when respect for my rational nature is their motive. But none of us can consistently will this. For, each of us can plausibly imagine ourselves coming into a situation where we are in a great deal of pain but our only hope for relief rests with agents who *are not* motivated by respect for rational nature. (Their motives might be more along the lines of a classical utilitarian's, for example.) Because each of us wills to have our well-being promoted even in such cases, none of us can at the same time will for all agents to always accord with M2a. So, according to FUL, none of us should act on M2a.

The reasons a human agent might have for discriminating against non-agents (as far as the promotion of well-being is concerned) could take a rather different form. For example, perhaps one of the most common reasons for refusing to make sacrifices for the sake of non-agents' well-being is that they are so rarely able to (deliberately) reciprocate;

and they are rarely able to harm us for refusing to assist them. But consider this maxim (M2b): In order to most efficiently pursue my ends, I will not go out of my way to promote the well-being of sentient beings that are not able to punish me or reciprocate. None of us can consistently will this as a universal law. For, each of us can realistically imagine ourselves coming into a situation in which (a) we are in a great deal of pain, (b) there are agents available who could alleviate this through permissible means and with very little effort or sacrifice, (c) they are our only hope for relief, and (d) we would be unable to repay or punish any of them. Each of us wills that we would have our well-being promoted in such a case. But if M2b were a universal law, the putative benefactors in such cases would discriminate against *us*. So, we cannot consistently will M2b as a universal law. According to FUL, acting in accord with this discriminatory maxim is morally impermissible; those who have adopted it should disabuse themselves of it.

Another common reason for discriminating against sentient non-agents is that it is rarely the case that any *agents* would punish one for it or reward one for *not* discriminating. But an argument much like the above can easily be constructed against the following maxim (M2c): In order to most efficiently pursue my ends, I will not go out of my way to promote the well-being of sentient beings except when agents would reward me for doing so or punish me for not doing so. For, each of us can realistically imagine ourselves in a situation where (a), (b), and (c) above are all satisfied, but (d*) *no one* would be able to reward or punish the putative benefactors. Each of us wills that we would have our well-being promoted in that kind of case, too, but with M2c as a

universal law there would be no hope of this. So, according to FUL, M2c is also off-limits for human agents.

M2a, M2b, and M2c share a common problem, namely, that no human agent can consistently will for other agents' treatment of *her* to always accord with any of these maxims (since this would, in some plausible situations, conflict with something else that she invariably wills). But the above examples are really just illustrations of a very general point which rules out rafts of discriminatory maxims. It is this: If you cannot consistently will for other agents' treatment of you to always be in accord with some maxim M, you cannot consistently will *for all agents to always act in accord with M*—in which case you violate FUL by acting on M. So, if an agent cannot consistently will for others to always act in accord with M in their treatment of *her*, she violates FUL whenever she acts on M. In this way, many of the common reasons behind most kinds of discrimination—whether against non-agents or against certain agents—can be seen as reasons on which no human agent should act.

One question now, therefore, is whether there is any maxim M that prescribes discriminating against non-agents (as far as the promotion of well-being is concerned) but which a human agent *could* will for all agents to accord with in their treatment of her. But refuting my above prima facie case requires more than coming up with such a maxim. The further question is whether those who *can* will for other agents to always accord with M also can sincerely claim that M captures *their genuine reasons* for discriminating against non-agents. For, otherwise, we are not talking about *their*

discriminatory maxims in the first place. It would be pointless to show that some human agents could will M as a universal law, if in fact none of those agents *has* M as a maxim.

So, a very special kind of maxim is needed in order to refute the prima facie case, that is, to show that some human agents who *do not* aim at promoting the well-being of non-agents are nonetheless in accord with FUL. It is not enough to come up with a discriminatory maxim that they could will as a universal law. It must be a maxim that at least one of these agents *actually acts on*. So, whatever maxim we might entertain, it would need to be shown that some human agents who actually act on it can also will to always be treated in accord with it. To be sure, I cannot survey all of the maxims that might be suggested, and I have no argument to the effect that the search is *bound* to be futile. But we can safely place the burden of proof on those who suspect that such a maxim is discoverable.

IV. HUMAN VIRTUE AND CONCERN FOR NON-AGENTS

Let me now connect the above results to my account of virtue. Recall, I have been claiming that the cornerstone of virtue is the moral commitment, that is, the stable, non-instrumental aim of living in accord with practical laws, or, practical principles valid for all possible agents. My primary goal in this chapter was to argue that, for human agents, living in accord with the practical laws entails aiming to benefit non-agents as well as agents. But in the course of doing so, I hope to have shown that there are problems with Kant's understanding of the practical laws themselves. By the end of Section II, we saw

reasons for doubting that the Formula of Humanity (FH) and the Formula of Universal Law (FUL) are one and the same, as Kant claims. Although rational nature is clearly the ground of FH, I argued that it cannot ground FUL. We get further confirmation of these principles' distinctness by combining the results of Sections I and III. In Section I we saw that the direct duty to promote other rational beings' well-being cannot be derived from FH. But we have just seen that a direct duty concerning the well-being of sentient beings *in general* can be derived from FUL. So, FH and FUL differ not only in their respective grounds but also in the duties they can generate.

I am inclined to say that we do not need to posit FH as a practical law in the first place. It seems to me that FUL can generate any duty we could correctly derive from FH. The arguments in the previous section could be adjusted to show, for example, that a maxim to treat others' rational nature merely as a means for one's own purposes cannot consistently be willed as a universal law by any human agent. Of course, FUL says nothing about what is and is not *an end in itself*. And the idea of rational nature as the sole end in itself has been taken as the foundation of Kantian morality. As we saw in Section II, however, Kant's arguments for this idea are unsuccessful. Moreover, it is unclear whether any argument can show that rational nature has the absolute value that Kant attributes to it. And if rational nature *is not* absolutely valuable, then FH, as stated, is not a practical law.

Part of the appeal of FUL is that it enables us to establish direct duties—again, including duties concerning others' rational nature—without claiming *any* bearers of

absolute or intrinsic value. The things an agent *in fact* values, or, wills to achieve, receive, or have promoted, place constraints on what kinds of maxims she can consistently will as universal laws. Precisely because of this, accord with FUL requires agents to adopt other-regarding aims that correspond with the things they will from other agents—regardless of whether any or all of these things are valuable in some objective sense.

So, we do not need to claim that well-being is objectively valuable in order to say that the (more or less) virtuous human agent aims to promote the well-being of sentient beings in general for its own sake. It would be sufficient if it could be shown that such an agent must have this aim in order to be in accord with FUL. But, like every other human agent, the virtuous human agent wills that, at least in certain situations, others would promote her well-being. So the *prima facie* case of the previous section is in play here as well. And defeating it is especially difficult once we restrict our attention to virtuous human agents. For, if such an agent is to discriminate against non-agents when it comes to promoting well-being, she would have to have the *right* reasons for doing so. And I think we can now see that these reasons are not to be found in Kant.

If my arguments in Section I are correct, her reason cannot be that promoting others' well-being is worthwhile only as a constituent of treating rational nature as an end in itself. I argued that treating rational nature as an end in itself does not entail promoting agents' agreeableness (pleasure) or preventing their suffering, which means that the right reasons for promoting *agents'* well-being would have to come from elsewhere. Might

she think that promoting others' well-being is worthwhile only as a *means* for protecting or enhancing others' rational capacities? In order to think this, she would have to believe that all of her other-regarding duties stem from the basic requirement to protect/enhance others' rational capacities. But I doubt there is any good reason to believe this. As we saw in Section II, Kant's reasons are deeply problematic.

On the other hand, Kant's FUL speaks *against* the more common kinds of reasons for discriminating against sentient non-agents, as we saw in Section III. Intuitively, we would think that a virtuous human agent *would not* discriminate against others on such grounds, e.g. that there would be nothing in it for her if she did promote their well-being, or that she can get away with refusing to promote it. I think most of us would agree that the uncoerced willingness to help others when there is no expectation of a return is a *mark* of virtue. I basically showed that FUL legitimates such judgments. But such judgments are no less legitimate when the "others" in question are sentient non-agents.

REFERENCES

- Carruthers, Peter, *The Animals Issue* (New York: Cambridge University Press, 1992).
- Copp, David & Sobel, David, "Morality and Virtue: An Assessment of Some Recent Work in Virtue Ethics," *Ethics*, 114 (April 2004), 514-554.
- Denis, Lara, "Kant on the Perfection of Others" (*Southern Journal of Philosophy*, 1999), 21-41.
- "Kant's Conception of Duties Regarding Animals: Reconstruction and Reconsideration," *History of Philosophy Quarterly* 17 (2000), 405-23.
- Driver, Julia, *Uneasy Virtue* (Cambridge: Cambridge University Press, 2001).
- Egonson, Dan, "Kant's Vegetarianism," *Journal of Value Inquiry* 31 (1997), 473-83.
- Foot, Philippa, *Natural Goodness* (Oxford: Oxford University Press, 2001).
- Gregor, Mary (ed. and trans.), *Practical Philosophy* (Cambridge: Cambridge University Press, 1996).
- Guyer, Paul, *Kant on Freedom, Law, and Happiness* (Cambridge: Cambridge University Press, 2000).
- Herman, Barbara, *The Practice of Moral Judgment* (Harvard University Press, 1993).
- Hurka, Thomas, *Virtue, Vice, and Value* (Oxford: Oxford University Press, 2000).
- Hursthouse, Rosalind, *On Virtue Ethics* (Oxford: Oxford University Press, 1999).
- Johnson, Robert, "Expressing a Good Will: Kant on Moral Worth", *Southern Journal of Philosophy*, Summer, 34:2 (1996), 147-168.
- "Happiness as a Natural End", in M. Timmons, ed., *Kant's Metaphysics of Morals: Interpretive Essays* (New York: Oxford University Press, 2002), 317-30.
- Korsgaard, Christine, "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations," *Philosophical Perspectives* 6 (*Ethics*, 1992), 305-332.
- The Sources of Normativity* (Cambridge: Cambridge University Press, 1996).

- Louden, Robert, "Kant's Virtue Ethics," *Philosophy*, 61 (Oct., 1986), 473-89.
- Martin, Adrienne, "How to Argue for the Value of Humanity," *Pacific Philosophical Quarterly* 87 (2006), 96-125.
- Meredith, James (trans.), *Critique of Judgment* (New York: Oxford University Press, 1952).
- Nozick, Robert, *Anarchy, State, and Utopia* (New York: Basic Books, 1974).
- O'Neill, Onora, "Kant on Duties Regarding Nonrational Nature II," *Proceedings of the Aristotelian Society* Supplement 72 (1998), 211-28.
- Sherman, Nancy, *Making a Necessity of Virtue: Aristotle and Kant on Virtue* (Cambridge: Cambridge University Press, 1997).
- Skidmore, James, "Duties to Animals: The Failure of Kant's Moral Theory", *The Journal of Value Inquiry* 35 (2001), 541-59.
- Slote, Michael, *Morals from Motives* (Oxford: Oxford University Press, 2001).
- Swanton, Christine, *Virtue Ethics* (Oxford: Oxford University Press, 2003).
- Wood, Allen, *Kant's Moral Religion* (Ithaca, New York: Cornell University Press, 1970).
- & di Giovanni, George (eds. And trans.), *Religion within the Boundaries of Mere Reason* (Cambridge: Cambridge University Press, 1998).
- "Kant on Duties Regarding Nonrational Nature," *Proceedings of the Aristotelian Society* Supplement 72 (1998), 189-210.
- Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999).