

# The Value of Humanity

Lisa Nandi Theunissen

Submitted in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy  
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

© 2012  
Lisa Nandi Theunissen  
All rights reserved

## ABSTRACT

### The Value of Humanity

Lisa Nandi Theunissen

My dissertation is on foundational questions about the value of human beings. This is a Kantian topic but I develop a proposal in a non-Kantian framework. I argue that to be a Kantian in ethics is to be committed to rationalism, but that the foundations of ethics should take account of the nature of human beings and our circumstances in the world. I develop a non-Kantian theory in which the value of human beings is no different, metaphysically speaking, from the value of other valuable things. Human beings have value, just as anything of value has value: because we are capable of being good-for something or someone. Most fundamentally, I argue that we are capable of being good-for ourselves. I propose that human beings have value in virtue of a capacity for having final ends, and that the capacity for having final ends makes us valuable because it makes us capable of living a good life, a life that is valuable because it is good-for the person who leads it. I show how the value of human beings gives everyone reason to treat human beings in certain ways. In particular, I show how everyone has reason not to destroy the capacity of human beings to have final ends, and, more positively, to help others realise their ends.

# Contents

<b>Acknowledgments</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Kant on Ends-In-Themselves . . . . .	3
1.2 A Normative Doctrine . . . . .	5
1.3 Kant on the Absolute Value of Humanity . . . . .	7
1.4 A Non-Kantian Account of the Value of Human Beings . . . . .	11
1.5 Open Questions . . . . .	14
<b>2 Kant's Commitment to Metaphysics of Morals</b>	<b>18</b>
2.1 Introduction . . . . .	18
2.2 Metaphysics of morals . . . . .	24
2.3 Kant's Argument for Metaphysics of Morals . . . . .	27
2.4 The Positive Role for Empirical Considerations . . . . .	29
2.5 Virtue-Ethical Lessons from <i>The Metaphysics of Morals</i> . . . . .	33
2.6 A Foundational Role for Human Nature? . . . . .	36
2.7 Bindingness Without Metaphysics? . . . . .	39
2.8 Conclusion . . . . .	44
<b>3 Responsibility and the Value of Intelligible Beings</b>	<b>46</b>
3.1 Introduction . . . . .	46
3.1.1 Regarding Others as Intelligible Beings . . . . .	47
3.1.2 A Dilemma About Responsibility . . . . .	48
3.1.3 Kant's Dilemma . . . . .	49
3.2 Velleman on Valuing Others as Intelligible Beings . . . . .	52

3.2.1	The Pre-existing Value of Intelligible Beings . . . . .	52
3.2.2	Right Appreciation: Respect and Love . . . . .	53
3.2.3	Exclusion of the Bad in Sensible Nature . . . . .	54
3.2.4	Love as an Ideal for Interpersonal Relations . . . . .	57
3.2.5	A Model of Interpersonal Relations . . . . .	58
3.2.6	The First Horn of the Dilemma . . . . .	60
3.3	Korsgaard on Respecting Others as Agents . . . . .	63
3.3.1	The Second Horn of the Dilemma . . . . .	66
3.4	A Disjunctive Conclusion . . . . .	72
3.4.1	Concessive Kantianism . . . . .	72
3.4.2	Rejection of the Distinction Between Intelligibility and Sensibility	74
<b>4</b>	<b>On Regress Arguments for the Value of Valuers</b>	<b>78</b>
4.1	Introduction . . . . .	78
4.2	The Notion of Value Simpliciter . . . . .	80
4.3	The Regress Schema: A Source Intuition . . . . .	87
4.3.1	Stage 1 . . . . .	87
4.3.2	Stage 2 . . . . .	88
4.3.3	Responses to the argument . . . . .	88
4.4	The Regress Schema: An End to the Chain of Dependence . . . . .	89
4.4.1	Stage 1 . . . . .	90
4.4.2	Stage 2 . . . . .	90
4.4.3	Two notes about Stage 1 . . . . .	91
4.5	The Alternatives . . . . .	94
4.6	Being valuable for Ourselves . . . . .	100

4.7	Conclusion . . . . .	101
<b>5</b>	<b>On the Value of Human Beings</b>	<b>102</b>
5.1	Introduction . . . . .	102
5.2	The Basis of Human Value . . . . .	106
5.3	The Explanation of Human Value . . . . .	113
5.3.1	The value of human beings as bearers of final ends . . . . .	116
5.4	The Response to Human Value . . . . .	121
5.5	The Responses We Should Have . . . . .	125
5.6	Conclusion . . . . .	128
	<b>References</b>	<b>130</b>

## Acknowledgments

Many people have been involved in shaping the direction of my project. I acknowledge various people in footnotes in various chapters, but I would like to acknowledge several people more fully here.

I am grateful to Jens Timmermann on the Kant side of things. Jens invited me to participate in a conference on Kant and the dignity of humanity at the University of St Andrews in May 2011, and it was the occasion of several important conversations about Kant on value. I am also grateful to Pat Kitcher for providing helpful feedback on the Kant side of things. I was very grateful to Pat for meeting me on the eve of 2012—while everybody else was beginning their New Year celebrations—before I was to leave to give a talk on Kant on metaphysics of morals.

I was fortunate to be able to attend David Velleman's undergraduate lectures on Kant's ethics, and to learn first hand from his distinctive take on Kant. As I remember, there were obstacles to entering the building at New York University where David was lecturing, and he was kind enough to meet me in front of the building in the mornings, so I could gain entry. Our brief exchanges through the corridors and up the elevator were the start of many subsequent conversations. David has patiently made his way through each of the chapters of my dissertation, and I am very grateful for his feedback on all of them.

Joseph Raz's influence on the direction of my thinking will be obvious. Much of my dissertation engages in a detailed way with Joseph's writing. Our working relationship has been (relative to my life as a philosopher) rather long. I was a nervous student in two of his seminars at the Law School at Columbia, and a nervous student in several Oxford-style meetings about this or that paper, where I had to defend something I had provisionally put together. Since then Joseph has become a mentor and friend. His guidance has, and will

continue to have, a lasting impact on my life. Philosophy would not be what it is for me without him.

It is not easy to put my gratitude to Katja Vogt into words, or to trace all aspects of her guidance on my work. This project grew out of a seminar Katja gave on Kant and his critics in the Spring of 2008. That seminar profoundly shaped my thinking about ethics, and about philosophy. It was also in that seminar that I began to learn how to read difficult texts in the history of the tradition. I have discussed every aspect of my project with Katja over many years. It was Katja who first challenged me to think about whether the notion of absolute value is a notion we fully understand, and she who brought me to think about the ancient view, in a Stoic formulation, that the good *is* benefit. But I have learnt from Katja in more indirect ways, too. I have attended nearly all her seminars at Columbia, and been a TA for several of her undergraduate courses. Without Katja, if I can put it this way, I believe I would be somebody else.

A person whose influence will be less obvious on the page is Bradley Weslake, but his influence has been profound. I met Brad when I was doing my Honours in philosophy at the University of Sydney in 2003. He shook up everything I thought philosophy was, often to my irritation. Our partnership over the eight years that followed meant that I engaged philosophically every day with someone who had some years experience over me, and a stronger hand. He has been an editor, an interlocutor, a friend, and a partner.



## I Introduction

My dissertation begins with the question of why human beings are to be viewed, treated, and generally related to in very special ways. It asks: what kind of value grounds this distinctive status? Accordingly, the starting-point of my project is the central ethical question of what we owe to others. I am concerned with our ethical relations with human beings, whose value I take to be distinctive, while allowing that there could be an analogous (if less ambitious) picture about what we owe to non-human animals.

From what we owe to human beings it is a short step to the *value* of human beings. We take that step when we ask *what it is* about human beings that makes us subjects of ethical concern. That we ‘have an elbow’, or ‘a vestigial appendage’, that we ‘walk upright’, are so many joking candidates.<sup>1</sup> But they make a serious point. If human beings are such as to merit being treated in certain ways, it is because something about us is *valuable*. The question of my dissertation is: what makes human beings valuable, and accordingly, such as to be treated in certain ways and not in others?

Much of the literature on the value of human beings begins with Kant. Recent emphasis on the so-called ‘Formula of Humanity’ version of the Categorical Imperative has made Kant the central interlocutor in accounts of human value generally. Now there are various ways of appealing to Kant on this topic, and of working out whether to construct a positive account of the value of human beings in a Kantian framework. A common approach is to engage closely with Kant and Kantian views on the value of humanity. It is not the approach taken in the present work. My approach is rather to ask, more basically, what is it to be a Kantian in ethics? What commitments make it the case that one is or is not a Kantian? With that in view, the thought is, we have grounds for determining whether to

---

<sup>1</sup>They are joking candidates of Robert Nozick.

construct an account of human value in a Kantian framework.

In Chapter 1, 'Kant's Commitment to Metaphysics of Morals', I argue that a definitive feature of Kant's moral philosophy is its rationalism. Kant insists that moral theory, at least at its foundation, cannot take account of empirical facts about human beings and their circumstances in the world. This is the core of Kant's commitment to 'metaphysics of morals', and it is what he sees as his greatest contribution to moral philosophy. The paper clarifies what it means to be committed to metaphysics of morals, why Kant is committed to it, and where he thinks empirical considerations may enter moral theory. The paper examines recent work of contemporary Kantians (Barbara Herman, Allen Wood and Christine Korsgaard) who argue that there is a central role for empirical considerations in Kant's moral theory. Either these theorists interpret Kant himself as permitting empirical considerations to enter moral theory, or they propose to extend Kant's theory so as to allow them to enter. I argue that these interpretive trends are not supported by the texts, and that the proposed extensions are not plausibly Kantian. Kant's insistence on the exclusion of empirical considerations from the foundations of moral theory is not an incidental feature of his thought which might be modified while the rest remains unchanged. Rather, it is the very center of his endeavors in moral philosophy. If we disagree with it, I argue, we have grounds for moving to a distinctly different theoretical framework.

In Chapter 2, 'Responsibility and the Value of Intelligible Beings', I show why Kant's rationalism may be thought to raise deep concerns. Kant has long been thought committed to a dilemma about responsibility: either we are never responsible for morally bad actions, or we are always responsible for morally bad actions. We are never responsible for bad actions because, for Kant, we are only fully agents when we act as we morally should, not when we fail to do so. We are always responsible for bad actions, for Kant, because we are always capable of moral agency, and therefore always responsible for failing to live up to

it. I consider whether either horn of the dilemma may be seized by appealing to lines of argument in David Velleman and Christine Korsgaard. Ultimately, I conclude that these arguments fail. The dilemma is generated by Kant's non-empirical conception of agency, and we should reject this conception. While it is true that we are rational agents, I argue that our empirical nature is also taken, and is right to be taken, as practically relevant. We relate to one another as rational beings, but also as beings with failings and vulnerabilities.

In sum, the first two chapters of my dissertation make the case that Kant is a rationalist in ethics, and that rationalism is a position we have reason to reject. Taken together, these chapters give my grounds for developing an account of the value of human beings in a non-Kantian framework.

But, of course, it is important to have Kant's account of the value of humanity in view. In many ways, Kant has set the terms—human beings as 'ends-in-themselves', human beings as 'absolutely' valuable—for subsequent discussions. By way of introduction, then, allow me to say a word about Kant's views, as I read him, on the value of humanity. It will become clear that Kant's treatment sets the terms for my own in several respects.

### 1.1 Kant on Ends-In-Themselves

Kant's claim that human beings are 'ends-in-themselves' is widely invoked, and it has achieved something of the status of a platitude. But it is worth pausing to ask whether it is a notion we fully understand. For despite the familiarity it has achieved, the idea that human beings are *ends* might strike us as rather puzzling.

As it emerges in Kant, the notion of an end-in-itself figures in what we might think of as a theory of action. Kant's is less a theory of action in general than a theory of those actions we *should* be doing. Given Kant's peculiar conception of normativity, his is an account of action in accordance with *laws* of practical reason. There are broadly two kinds of law

of practical reason, the moral law and the law (or more weakly, the rule) of instrumental reason. The talk of ends comes as Kant is outlining what is involved in the capacity to act in accordance with the representation of laws (4:412). Like Aristotle, Kant thinks that action is for the sake of ends. This is a moral-psychological point about that in view of which we determine ourselves to act. There are the ends which vary among us, and there is happiness as the end common to us all. Unlike Aristotle, however, Kant thinks there is something distinctly moral in view of which we determine ourselves to act. Kant calls this an end-in-itself. The end-in-itself is that for the sake of which we act when we act morally.

For Kant, human beings, as rational beings, meet the criteria for being ends-in-themselves. On the face of it, human beings ought to strike us as surprising candidates for being ends of any kind. What is puzzling about thinking of human beings as *ends* is that an end is to be brought about; it is an object of choice. And it is not clear what it would mean to say that human beings are to be brought about. Kant handles this infelicity by having ends-in-themselves be different *kinds of end*: independently- or self-existent ends (4:437). The thought is that human beings are ends, not in the sense that they are to be brought about as effects of our action, but in the sense that they are constraints or limits on action. We might say that while Kant takes over Aristotle's teleological conception of action in taking action to be for the sake of ends, the notion of an end-in-itself is robbed of its teleological import.

Still, I take it human beings may yet be said to qualify as 'ends' in at least two ways. First, an end may be parsed as that for the sake of which we act, and human beings may intelligibly be said to be that for whose sake we act. So we can intelligibly say things like: "I brought the meeting to a close for her sake" or, "You did not give back the money for his sake, but for your own". Second, I take it the talk of ends is meant to exploit a contrast with means, so as to make a normative point. The notion of an end-in-itself picks out that

in a theory of action which cannot figure merely as a means to something else, and in that sense figures always as an end. While activities like making tea or fetching children are ends from one point of view—ends from the point of view of boiling water or getting in the car—from another point of view they are means to further ends—to refreshment and having the children home. By contrast, the end-in-itself in Kant cannot lose its status as an end to become a mere means to something else. The qualification ‘mere’ here is important. For Kant, there are other ends with respect to which an end-in-itself may be taken as a means. So we might pay a taxi driver to take us home. Kant’s point is not that we cannot treat human beings as means to other ends we have. It is just that we should not treat them *merely* as means. It is in this sense that they cannot lose their status as ‘ends’. In these two respects, then, I take it that there is at least a diminished sense in which human beings may be thought of as ‘ends’.

Note that Kant also puts the difference between ordinary ends and ends-in-themselves in terms of a difference in our motivational orientation. Our motivation to pursue ordinary ends is a function of desire, while our motivation to pursue ends-in-themselves is a function of a rational motive. In this way, Kant’s account of what is involved in acting in accordance with the representation of practical laws has two components. There must be an end and there must be a source of motivation.<sup>2</sup> There are, then, two kinds of ends and two kinds of motivation, corresponding to the two kinds of practical law.

## 1.2 A Normative Doctrine

An end-in-itself, then, is that for the sake of which we act, when we act morally. It is an end we are to have irrespective of desire. And it is a distinctive kind of end in being such that

---

<sup>2</sup>Or so Kant lists the components when he outlines ‘the matter’ of practical laws, where he had hitherto outlined their ‘form’ (4:436-7).

it is to be acknowledged as a constraint on action rather than as something to be brought about. The account of human beings as ends-in-themselves is ultimately an account of what that acknowledgement comes to. It is an account of how we should behave with respect to human beings. Put colloquially, the doctrine says that we are to engage with human beings always in such a way that we recognise that they are capable of making up their own minds about things.<sup>3</sup> So Kant says we must treat human beings as limits on choice, and as objects of respect (4:428). Later he will add that we are to treat human beings in such a way that we consider their concerns, projects and interests—their *ends*—as things we must work to further (4:430). So what it means for human beings to be ends-in-themselves is for human beings to be treated in these particular ways.<sup>4</sup>

This being so, the doctrine of human beings as ends-in-themselves does not by itself tell us *why* human beings are to be valued as they are. The doctrine does not settle the question of whether human beings are to be treated as ends-in-themselves because they possess some special property, or some distinctive *kind of value*, or because ideally rational agents would so treat them; it does not tell us whether human beings are to be treated as ends because I would have others treat me that way, and on pain of inconsistency, must so treat others. To that extent, the doctrine may be thought to push our question back, our question being: what makes it the case that human beings are valuable, and accordingly, to be treated in certain favourable ways.

As we might put it, then, being an end-in-itself is a second-order property. To be an end-in-itself is to be the kind of thing that is to be treated in certain ways—as a limit

---

<sup>3</sup>While we might gloss this by saying that we should treat human beings as autonomous, really autonomy should be understood as meaning, more strictly, capable of legislating, and governing oneself with a view to, the moral law. Or so the emphasis comes to be laid in the so-called Autonomy Formula of the moral law. See 4:431; 4:435-4:436.s

<sup>4</sup>See O'Neill (1985).

on choice, an object of respect, and something whose ends are to be furthered. Now the expression ‘ends-in-themselves’ is sometimes used to denote that in virtue of which human beings are to be treated in certain ways. That is, the expression is sometimes used to denote a first-order property, a property which features in an explanation of the special treatment of human beings. So it is said that human beings must be treated with respect *because* they are ends-in-themselves. If what I have said is right, however, that would be to say no more than that human beings must be treated with respect because human beings must be treated with respect. I would add that this purported explanatory use of the phrase stretches the notion of an ‘end’ quite beyond the extended sense granted above, the sense granted for ‘independent-’ or ‘self-existent’ end. For what does it mean to say that an object has the property of being an end (or a means for that matter)? It is agents who take objects as ends or means; objects by themselves cannot bear these relations. The notion of an end must be relativised to an (possible) agent, or valuer.<sup>5</sup>

### 1.3 Kant on the Absolute Value of Humanity

Why does Kant think human beings must be treated as ends-in-themselves? As I read him, Kant’s explanation is that human beings have ‘absolute value’. Absolute value is said to be something the existence of which in itself has value (4:428), or *inner* worth (4:435). The bearer of absolute value, as Kant puts it figuratively, is like a jewel that shines by itself, having its full worth in itself (4:394). While it involves other notions besides, the notion of absolute value involves at least the idea of value that does not derive from its relationship to anything else.

---

<sup>5</sup>That is, presuming we do not hold a teleological world view. For Aristotle it makes sense to say that ends are goods and goods are ends. But we no longer take this kind of equation to be plausible—we don’t think of things as themselves end-like.

The notion of absolute value is contrasted with relative value (4:428). What has relative value may be priced and admits of comparison, replacement and equivalents (4:435). The explanation of the value of relatively valuable things—of everything that is not absolutely valuable, for these categories are taken to be exhaustive—is that we are inclined towards them. Kant does not have a sophisticated theory of relative value, and he seems to take inclination to be a basic explanatory concept, so that the concepts of interest, need and benefit are reducible to it.<sup>6</sup> This is questionable since we may well have no inclination towards what may be of benefit to us, as in, a trip to the dentist. But the category of relative value is seemingly meant to include what has worth because we want it, because it makes our lives go well, meets our needs, or is of benefit. About all values of this kind Kant is a kind of projectivist, so that our wanting (or needing, or being benefited by) something is not a response to a seeming good-making property in it, but is conferred by desire.<sup>7</sup> Again, while it is characterised in other terms besides, the notion of relative value involves at least the idea of value that derives from its relationship to something else, as in Kant's account, to the inclinations of human valuers.

Kant's argument for the existence of absolute value, and for human beings as bearers of absolute value, proceeds in two stages. There must be something of absolute value for without it there could be no moral law (stage 1); human beings meet the criteria for having absolute value (stage 2). We have already met stage 1 of the argument, but we can restate it intuitively as follows. The moral law is categorically binding on human beings—it holds

---

<sup>6</sup>He writes, for example, of “inclinations and the needs based on them” (4:428). Later (at 4:435), Kant draws a distinction between value that is relative to inclination, and value that is relative to a certain taste, or delight. So there are in fact two basic concepts at work here, inclination and taste. They correspond to what has market price and fancy price, respectively (4:434-5).

<sup>7</sup>Kant sometimes puts the relative value idea in terms of what has conditional value (4:428), where that means value conditional on our inclinations and taste, but also, if we turn back to the opening of the *Groundwork*, value conditional on circumstance. Relative values are such that they may be good in some circumstances, and bad in others. They are not, as Kant also puts it here, good without limitation (4:393-4).



irrespective of inclination. Action in accordance with the representation of the moral law requires an end, or good. Since the law is categorically binding, this end or good cannot be relative to inclination—it cannot be good because we want it, and not otherwise. It must be an end or good that simply has value—has absolute value. If there is a moral law, there is absolute value.<sup>8</sup> The second stage of Kant’s argument, such as it is, is that, unlike the other candidates he considers—the objects of our inclinations, inclinations themselves, non-rational beings—the nature of human beings, as rational beings, marks them out as absolutely valuable (4:428). Human beings are absolutely valuable on account of their nature as rational beings.<sup>9</sup>

So stated, what are we to make of Kant’s explanation of the distinctive value of human beings? We can take hold of Kant’s argument at both stages 1 and 2. Of stage 1, we might deny that moral requirements must be grounded in values.<sup>10</sup> Or we might agree that moral requirements must be grounded in values, but deny that moral requirements are of a sort (such as to be categorically binding) that they must be grounded in values that are likewise absolute. I am sympathetic to this second line. To look ahead, my scepticism about

---

<sup>8</sup>Note that Kant’s argument is hypothetical. He is not here arguing that there is a moral law, he is so far only outlining what the law would need to be, and what the capacity to act in accordance with it would need to involve. As Kant puts the argument: there is something of absolute value because “without it nothing of *absolute worth* would be found anywhere; but if all worth were conditional and therefore contingent, then no supreme principle for reason could be found anywhere” (4:428). The task of showing that there is a moral law is the burden of Section III of the *Groundwork*.

<sup>9</sup>In stating the two stages of the argument this way, I agree with the interpretations given in Timmermann (2006) and Vogt (2009). I disagree with the interpretations given in Korsgaard (1986a) and Wood (1998), and with the general line of interpretation to which they have given rise. Among recent defenders of the Korsgaard line, see Martin (2006) and Sussman (2003). Among recent detractors of Kant read along Korsgaardian lines, see Regan (2002).

<sup>10</sup>Important interpretive questions come up here about, as it is sometimes put, deontological versus teleological readings of Kant. See for example the lines taken in Herman (1993a), and Johnson (2007). See also Timmermann (2006). My own view is that the moral law is grounded in a value, for Kant, but in such a way that the value turns out to be a postulate. For that reason, Kant is not what we would think of as a value realist.

absolute value, or in my terms, value simpliciter, is related to a scepticism about categorical bindingness.

Of stage 2, consider that, as Kant goes on to tell us what it is about rational nature that makes it so valuable, the emphasis comes to be laid on our capacity to be moral. As the famous passage has it: “Hence morality, and humanity insofar as it is capable of morality, is that which alone has dignity”, where to have dignity is to have absolute value (4:435). I admit I find this to be an overly narrow, or overly moralised, conception of what makes human beings valuable. What is more, at times Kant writes as if it is not the *capacity* to be moral that makes human beings valuable, but the state of *having* a good will. As Kant puts it in one of a number of similar passages, it is the person who wills to act from basic principles that has dignity, and is the object of respect (4:435). So while Kant has long been held to be the Enlightenment proponent of the view that all human beings have distinctive value just in virtue of being a certain kind of being, when we look closely at the texts, he seems to say that more—that *merit*—is needed. That is, it looks as if Kant is saying that only morally good human beings have the relevant kind of value.<sup>11</sup>

Finally, there is the possibility not only that not *all* human beings have absolute value for Kant, but that *no* human beings do. This is a possibility not just for the reason that likely no one has a good will. For there are passages which suggest that the proper object of respect, the mode of valuation proper to objects of absolute value, is not human beings in the first place, but the moral law. Kant says that “Any respect for a person is properly

---

<sup>11</sup>Darwall (2008) considers the challenge posed by these passages to the orthodox, non-merit, reading of Kant on the value of humanity, though he ultimately concludes in favour of the orthodox view. Dean (2006) takes a different position. He defends the view that only human beings with a good will have absolute value for Kant, mainly for the reason that the *Groundwork* opens with the claim that only the good will has absolute value. Vogt (2009) suggests we think of Kant as going back and forth on this question. She argues that while Kant never gives up the idea that human beings simply have absolute value, he shifts emphasis in the *Metaphysics of Morals*. There the emphasis is less on treating human beings as objects of value, and more on treating the values (the ends) of human beings as the objects of concern, and the locus of action.

only respect for the law [...] of which he gives us an example” (4:401 n). In these moments, it looks as if the value of human beings is relational after all: the value of human beings depends on their standing in an appropriate relationship to the moral law. Now we might say that Kant thinks that it is in the nature of human beings to be moral, or to be capable of being moral, so that the relationship of human beings to the moral law is not a contingent one. But even if it is constitutive of human beings that they be related in some way to the moral law, it is still the moral law that has non-relational value, human beings only derivatively so.

#### 1.4 A Non-Kantian Account of the Value of Human Beings

My own account of the value of human beings in a way begins by asking Kant’s question: are human beings absolutely valuable? The notion of absolute value, as it has been explicated above, is a notion of non-relational value, of something’s being valuable independently of its relationship to anything or anyone. In fact, my interest is slightly more focussed. My interest is in whether human beings are valuable simpliciter, that is, valuable independently of being good-for anything or anyone—independently of being useful for something, or beneficial in some broader way. This is an aspect of the notion of absolute value which interests Kant, as when he characterises the good will as being such that usefulness would add nothing to its worth, and fruitlessness take nothing away (4:494).

The distinction between good simpliciter and good-for is an *instance* of the distinction between non-relational and relational value. But unlike this second distinction, which is broad enough to encompass meta-ethical questions, the distinction between good simpliciter and good-for is a distinction in the metaphysics of value. As we might put it, it pertains to the structure of value. It is possible to have a range of meta-ethical theories of good simpliciter and of good-for, and the distinction cross cuts questions about, for example,

realism and anti-realism. In this work, I leave these meta-ethical questions to one side, while recognising that they would be part of a full account of the nature of the value of human beings.

It is something of a background hypothesis to my dissertation that there is no such thing as value simpliciter. My background hypothesis is that all value is value-for something or someone, and that the value of human beings is no exception. Now there are different ways to try to defend this hypothesis. One could construct a historical genealogy of the notion of value simpliciter, and show it to depend on assumptions we need not accept. For example, it may be shown to rest on theological assumptions.<sup>12</sup> Alternatively, one could argue that good *means* good-for.<sup>13</sup> Or one could argue that good simpliciter has no practical relevance for ethical life.<sup>14</sup> My own approach is to argue that the notion of value simpliciter is theoretically superfluous—that we can do without it. As a corollary of that, my focus is in large part on showing what an account which appeals only to the notion of value-for might look like, and how it can capture core desiderata for an account of the value of human beings. Chapters 3 and 4 of my dissertation are companion papers. They carry different burdens of this task, but the lines of argument are mutually dependent and supporting.

In Chapter 3, ‘On Regress Arguments for the Value of Valuers’, I examine a common argument schema for the value simpliciter of human beings. It begins with an argument for value simpliciter. There must be something of value simpliciter for anything to have value; the chains of dependence between values must come to an end. According to prominent versions of this argument, put forward by Joseph Raz and others, human beings meet the criteria for having value simpliciter. I reject this argument schema, and the notion of

---

<sup>12</sup>See Vogt (2009).

<sup>13</sup>On the semantic side of the question, see Geach (1956), Foot (1985), Thomson (2001, 2008).

<sup>14</sup>This is a way of putting the approach taken by Kraut (2011).

value simpliciter as theoretically superfluous. If we reject value simpliciter, the chains of dependence between values must be such that human beings are valuable either: (i) because we are good-for objects or activities of value; (ii) because we are good-for human beings (or other beings); or (iii) because we are good-for ourselves. I give reason to favour (iii). I argue that objects and activities of value have value because they have value-for human beings (or other beings), and that human beings have value because we are capable of standing in a relation to ourselves, a relation of being good-for ourselves. This makes the value of human beings non-derivative on the value of something else, but without invoking the notion of value simpliciter.

In Chapter 4, ‘On the Value of Human Beings’, I give a positive account of the value of human beings. I take the basis of human value to be that we are capable of having final ends, that is, interests, projects and relationships that are pursued for their own sake. By working through a proposal by Samuel Scheffler, I argue that to have a final end is (i) to believe that the end is valuable, (ii) to be guided by the end in long-range deliberation, (iii) to be engaged with the end in a sustained way over time, and (iv) to be emotionally susceptible to successes and failures in pursuit of the end. The capacity for having final ends makes us valuable, I argue, because it makes us capable of leading a good life. And a good life is valuable because it is valuable for the person whose life it is. The most fundamental explanation of the value of having final ends, then, is that it makes us (in this sense) good-for ourselves. The explanation has the air of a paradox. For one, it raises a question about the reasons we have to respond to human beings. If the capacity to have final ends makes us valuable because it makes us good-for ourselves, what reason should others have to respond to us? This is an instance of a more general worry, raised by Donald H. Regan, about the normative force of goodness-for. I argue that the worry rests on a misunderstanding about the theoretical role of goodness-for. That human beings are valuable because we are

good-for something or someone explains the value of human beings. That is, it explains how human beings genuinely have value—value which gives everyone reason to interact with human beings in specifiable ways. In particular, I argue that we have reason (i) not to destroy the capacity of human beings to pursue their own ends, and (ii) to help others pursue their ends.

My project ultimately relates to a wide range of issues in normative ethics, meta-ethics, and the metaphysics of value. Looking over it now I am struck above all by the open questions. I am struck too by the ways in which one's philosophical argumentation does not always secure the conclusions one had hoped for, or take one in the direction one anticipated. Still, my hope is to have laid the ground for an account of the value of human beings in which we are valuable, as other valuable things are valuable, because we are valuable-for something or someone. My hope is to have shown how this approach can capture important desiderata for an account of the value of human beings, for example, central obligations to treat one another in very particular ways. Much of what will need to be said is future work. In what remains of this introduction, I will describe some of questions, or projects, which naturally emerge from work in my dissertation.

## 1.5 Open Questions

While I lend support to the thesis that good is good-for in my dissertation, it remains to be made out what it is, meta-ethically speaking, for something to be good-for something or someone. I aim to develop a relational realist account of good-for. The first part of this label, 'relational', refers to the claim that good is good-for. The second part, 'realist', refers to the claim that, as I anticipate, there are frequently, though in some cases it may be indeterminate, objective facts of the matter about what is good-for something or someone. In my view, whether something is good-for something or someone depends on facts about

the nature of the subject, and facts about the valuable activity, object, or state of affairs in question. As I see it, then, we should not understand what is good-for someone to be a 'subjective' matter. And yet, I do think that a subject's interests, desires and preferences frequently bear on what is good-for them. This is not because what we want or prefer determines what is good-for us, but because it is often good-for subjects to have their interests, desires and preferences met (there are obvious exceptions!). There are further, and to my mind interesting, epistemological questions about how we can know that something is good for someone. Do we have to be brought up in a certain way? Is it something we can in some sense 'perceive', though we cannot fully explain why? But these are further matters.

A second topic to emerge from my dissertation is whether the pursuit of ethically bad ends is bad-for the person who pursues them—such that it will make her own life go badly. I take a stand on this question in my dissertation, suggesting that it will. This is an ambitious claim, and it is not one I adequately defend. Still, I find it plausible, and it is something I hope to take up in future work. One strategy is to argue that the notion of a good life for human beings is not reducible to a non-normative notion of biological functioning, or health. The thought would be that a person's life may be well going in material respects—the villain may be physically and psychologically healthy—though her life will not count as well-going in the relevant sense. The notion of a good-life, a life that is good-for the person who leads it, will be understood to be irreducibly normative. What does that mean? It means that the notion of a good life is not purely descriptive, but involves notions like 'well', 'flourishing' and so on—normative terms that cannot be eliminated. Then the argumentative burden will have shifted from something's being good simpliciter, to a life's being good-for the person who leads it, where the latter is not something we can further analyse in non-normative terms. Where would the theoretical gain lie, then? Is it not the case that the same, or analogous, metaphysical and epistemological worries

come up for the notion of a good life? One might think there is a clear epistemological gain in shifting from theorising about the good to theorising about the good-for human beings, since the notion of a good life (understood normatively) is an object of study in the non-fundamental sciences, and of humanistic reflection more broadly. The notion of a good life, though normative, may be thought a more tractable notion.

Another strategy, and one I am also sympathetic to, would be to deny that the villain, the one who pursues bad ends, is leading a good life in material respects—is functioning well. The strategy would be to turn the tables on the objector, and urge her to find a real case of a well-functioning villain.<sup>15</sup> The thought would be that, as a matter of empirical fact, villains tend to be psychologically and cognitively disordered. And there may be deep reasons for this, owing, for example, to our constitution as agents. I am sympathetic to the idea, raised for example by Thomas Nagel in *The View From Nowhere*, that we are normatively oriented towards the good. Or one might point to the sense in which human beings are social beings, that we exist in dependent relationships with one another, and that the life of the villain puts him deeply at odds with other people. I take it that the thought that it is the good person who has a good life is deeply Platonic (it is the view of the Plato of the Republic), and, like Plato, I think it is difficult to argue for. In my view, however, the burden is shifted away from an implausible conception of value (value as value simpliciter), and towards something that is, if difficult to prove, likely to be true! I should add that one can take the Platonic line while thinking that good-for is not something we can further analyse. We might say that someone whose life goes badly will exhibit certain features (lack of order, psychic conflict, impaired relationships with people, and so on). But you might think there will still be the question of what makes these features of a bad life? You might think this is the analogue of G. E. Moore's open question argument in the

---

<sup>15</sup>My thinking about this option stems from discussion with Katja Vogt.



domain of good-for. And I am sympathetic to some version of it.

A third topic is the relevance of so-called agent-relative considerations in ethics. It is sometimes argued that the notion of what is good-for someone is normatively inert, in the sense that the good-for someone gives rise to reasons for them, but not to others. If true, this would be a bad result for theories of the good-for human beings, and for prospects of making 'good for' central to ethics. I reject this argument in my dissertation. I argue that the fact that we explain the value of something in terms of its being good-for someone, does not affect the *scope* of the reasons to which it gives rise. As we might put it, good as good-for gives rise to perfectly 'agent-neutral' reasons. But this line of reply can seem to push matters too far in the other direction, in failing to capture intuitions about reasonable partiality. We tend to think that our own good, and the good of those close to us, is a matter of special importance to us. It guides our deliberations in ways that it does not guide the deliberations of strangers. The challenge is to find a way to accommodate these reasonable forms of partiality. I am interested in developing a proposal on which, while what is good-for someone gives rise to (at least 'pro tanto') reasons for everyone, we have instrumental reasons to attend to the good of those proximate to us. We have reason to attend to the good of those we are best equipped to attend to, given our competences, our knowledge of the facts, our standing with respect to them, and so on. This might seem to make the ground of our special obligations to ourselves and those close to us too weak, or of the wrong kind. To that I would argue that, while it is sufficient to capture important forms of partiality, the ground *is* weak. I would argue that this is so in virtue of our relatedness to all human beings, and the good of all human beings.

## 2 Kant's Commitment to Metaphysics of Morals

### 2.1 Introduction

Kant has long had an image as an austere and aprioristic moralist.<sup>16</sup> This is perhaps of little wonder given that Kant's avowed methodological commitment in moral philosophy, as he tells us in the preface to the *Groundwork*, is sharply to separate its rational from its empirical parts, and to work out the foundations of morality with the rational part alone. Since a notion of obligation lies at the heart of Kant's moral theory, Kant will put this as a commitment about the ground or basis of obligation. The ground of obligation is not to be sought in the nature of human beings or their circumstances in the world; the ground of obligation is to be sought a priori, in reason (GW 4:389).<sup>17</sup> Since obligation has a master principle, the Categorical Imperative, Kant will derive the master principle of morality, and from it, a system of duties, a priori. So much is the purview of moral philosophy proper, of what Kant calls 'metaphysics of morals', and Kant is doing metaphysics in each of his canonical practical works.<sup>18</sup> The remit of empirical ethics, what Kant calls 'practical anthropology', is to account for the material conditions which stand in the way of fulfilling obligations. But anthropology is separate from moral philosophy proper, and is given no systematic treatment by Kant himself.<sup>19</sup> This anyway is the orthodoxy. But the orthodox

---

<sup>16</sup>Bernard Williams has been one of the more insistent on this conception of Kant in contemporary discussions. See, for example, the opening of Williams (1976).

<sup>17</sup>References to all of Kant's works, except the *Critique of Pure Reason*, are to the volume and page numbers of the Preussische Akademie edition, and will be given in the body of the text. References to the *Critique of Pure Reason* will follow the standard method of citing page numbers in the first and second editions (A and B).

<sup>18</sup>By the canonical works I mean the *Groundwork*, the *Critique of Practical Reason*, and *The Metaphysics of Morals*.

<sup>19</sup>This is notwithstanding the publication of *Anthropology from a Pragmatic Point of View*, a record of Kant's notes for his paid lectures on anthropology in the form of a textbook. For an account of the provenance of the work, see Kuehn (2006, especially p. ix).

image of Kant has been shifting over the past thirty years or so, and a friendlier, a more humanistic, an altogether more empirically informed, Kant has grown up in its stead.

The shift is broadly coincident with emergent interest in Kant's late work, *The Metaphysics of Morals*.<sup>20</sup> While it goes over some of the pure foundations from the *Groundwork*, this book is more concretely interested in the form morality takes in human life. For example, where the *Groundwork* had given just a handful of applications of the principle of morality, *The Metaphysics of Morals* exploits very general facts about human beings to provide a fuller picture of our moral duties to ourselves and others. In the book's first part, The Doctrine of Right, the principle of morality is shown to yield externally enforceable laws for public life.<sup>21</sup> And in The Doctrine of Virtue there is discussion of various topics in moral psychology.<sup>22</sup> Included there, too, are casuistical asides on concrete matters such as the morality of serving wine at dinner parties (MM 6:428), or of signing a letter 'Your Obedient Servant' when the custom of politeness strains sincerity (MM 6:431). There is practical advice on the moral instruction of young people;<sup>23</sup> and practical concern for making morality effectual in the life of grownups.<sup>24</sup> *The Metaphysics of Morals* raises questions, in this way, about the status of applied and empirical considerations in Kant's late work. It

---

<sup>20</sup>At least in the anglophone world, that interest has coincided with Mary Gregor's 1991 translation, the first complete translation since the original English language publication in 1799. For notes on the history of the translation, see the Further Reading in Kant (1996b). For praise of Gregor's translation and a sense of its importance, see Dahlstrom (1993).

<sup>21</sup>As Sullivan (1996) notes in his Introduction to the Cambridge Edition, this is an important datum against the thought that the domain of Kantian morality is individual, personal life (p. xii). It may also bear on efforts to come to terms with the charge that Kant is socially disinterested, or a moral 'individualist'. Baier (1994) puts one version of that charge this way: "What is of interest for my purposes is his [Kant's] taking of social cooperation and group membership as not of primary moral importance" (p. 249).

<sup>22</sup>These include the relationship between pleasure, desire and the interests of reason (MM 6:211-14; 6:378), moral emotion (MM 6:399 ff.), and self-governance (MM 6:408 ff.).

<sup>23</sup>See Sections I and II of the Doctrine of the Methods of Ethics.

<sup>24</sup>See MM 6:397, 6:408, 6:411, 6:485.

has led some commentators to urge that Kant wholly reconfigures the *Groundwork* division between metaphysics and anthropology in his mature writings, so that there is development across the practical oeuvre.<sup>25</sup>

This developmental story finds new emphasis among commentators interested in the status of virtue ethical considerations in Kant.<sup>26</sup> These commentators argue that *The Metaphysics of Morals* makes room for notions of moral development, virtue, character, and moral perception. This has led to a reevaluation of the core doctrines standardly associated, for better or worse, with Kantian ethics. To objections against the Kantian model of deliberation—that it is rule-bound and inflexible; or against the ‘thinness’ of the Kantian person—that it is empirically insubstantial, the appeal to virtue in Kant has looked to yield ready lines of reply.<sup>27</sup> Indeed, the incorporation of virtue ethical considerations has been so pervasive that it has been possible to argue that there is nothing distinctive about virtue ethical, as against Kantian, theories in normative ethics.<sup>28</sup>

---

<sup>25</sup>See for example the remarks of Wood (1996, p. xxx).

<sup>26</sup>Among them are Herman (1993b, 2007), O’Neill (1984), Sherman (1997), Korsgaard (1996a), and Baron (1995).

<sup>27</sup>For an example of a Kantian response to the first kind of charge, see O’Neill (1984). For a forceful statement of the second kind of charge, see Williams (1976).

<sup>28</sup>This is a central thesis of Nussbaum (1999). Nussbaum argues that to treat virtue ethics as a distinct position from Kantianism (and utilitarianism) involves a category mistake, for all of these theories have substantive conceptions of virtue. Encapsulating many of the ideas in the above paragraph, Nussbaum writes:

In one way, this increasingly popular way of talking [of virtue ethics as a distinct position in ethics] is an obvious category mistake. Immanuel Kant has a theory of virtue, and devotes a good deal of attention to its exposition. Although *The Doctrine of Virtue* was at one time a relatively neglected part of Kant’s moral philosophy, read only by specialists, it is now widely discussed, and widely recognized as central. Nobody can any longer think of Kant’s view as obsessed with duty and principle to the exclusion of character-formation and the passions. [...] Moreover, the rediscovery of Kant’s theory of virtue has also led to serious reevaluation of the substantive positions of his other ethical writings, as scholars depict a Kant who is less rigorist and more flexible, less concerned with abstract principle and more concerned with the exercise of moral judgment, than the Kant of previous generations (p. 165).

At the same time, if from another direction, efforts have been made to fill out Kant's moral anthropology, the empirical part of moral philosophy to which he never gave systematic attention, and commentators have turned to *Anthropology from a Pragmatic Point of View*, the essays on history, and *Religion within the Limits of Reason Alone*.<sup>29</sup> On the basis of these texts some have claimed that far from being a subsidiary part of Kant's moral theory, empirical considerations about human nature actually lie at its foundation.<sup>30</sup> To the question, "What is Kant's ethics really about?", Allen Wood (1991) has urged that where others would say freedom, or the nature of practical reason, "I say that Kantian ethics is most fundamentally about human social conflict, its psychological roots and historical meaning" (p. 345). This has led to an air of sea-change, even revolution, in Kant interpretation, with commentators urging their peers to give up the aprioristic image of Kant as thoroughly stereotypical and interpretively unsound.<sup>31</sup>

My concern in what follows is to resist these recent trends. Kant's commitment to metaphysics of morals has been insufficiently appreciated in the literature. This is no doubt partly because Kant's ethics is largely treated independently from his metaphysics generally, with insufficient attention to their relationship.<sup>32</sup> But I will urge that Kant's commitment to

---

<sup>29</sup>See for example the essays on Kant's anthropology edited by Jacobs and Kain (2003), and the monograph by Loudon (2000).

<sup>30</sup>For example, Wood (1991) writes of the "anthropological foundations" of Kantian moral theory (p. 347), and of Kantian theory as "deriv[ing] its intellectual power mainly from its anthropological insights" (p. 345).

<sup>31</sup>On giving up stereotypes, see Wood (1999, p. 10); on the revolutionary gain in so doing, see Wood's description of the aim of his book: to "help to transform our conception of our own history and of ourselves as heirs of the Enlightenment. The aspiration of this book is to contribute in some small way to that revolution" (1999, p. 14).

<sup>32</sup>This has been remarked by Neiman (1994), who writes: "The excellent and growing body of recent work on Kant's moral philosophy has not devoted sufficient attention to the question why Kant holds that only reason can function as the source of moral principles" (p. 3). Neiman hypothesises that the lack of attention is due to treating Kant's moral philosophy separately from the critical philosophy. Foremost among commentators who have given systematic attention to Kant's notion of metaphysics of morals is O'Neill (1989). See also Gregor (1963, ch. 1). Note that while Gregor's work is cited as an authority by others—by, for

metaphysics of morals—to establishing the foundations of morality and its master principle a priori—sets real limitations on interpretation. The commitment is so central that to give up on it is to give up on the Kantian project. That is, I will argue that it is definitive of a Kantian position that it exclude empirical considerations from moral foundations.

My motivations here are both interpretive and philosophical (if you will allow the distinction). I would resist the idea that by incorporating empirical considerations, Kantian ethics is made more philosophically attractive.<sup>33</sup> Recent Kantians have sought to offset the idea that Kant's ethics is abstract,<sup>34</sup> ahistorical,<sup>35</sup> or unduly concerned with the motive of duty,<sup>36</sup> where the thought would be that a more concrete, historically situated, and sympathetically motivated theory would be preferable. But must we think this? We might make the general point that simplicity and abstractness are significant philosophical virtues, perhaps especially in light of the complexity of moral phenomena. We might enlist Kant's own observations about the importance of principles in light of our tendencies to rationalise

---

example, Louden (2000) whose opening chapter is on the topic—it is unfortunately out of print. Another resource on the topic is Beck (1960).

<sup>33</sup>I am grateful to Katja Vogt for discussion on this point.

<sup>34</sup>So Frierson (2003) writes in the Preface to his monograph:

From my earliest exposure to Kant's moral theory, I was drawn to his emphasis on the centrality of freedom but bothered by the apparent abstractness of the moral law. Thus when I first approached the *Anthropology* many years ago, I did so with excitement. I hoped that Kant would incorporate all the rich details of human life that I found lacking in his *Grounding* [...] At first the *Anthropology* seemed more amusing than philosophically satisfying. However, as I came to appreciate the details of Kant's *Anthropology* and as more neo-Kantians incorporated anthropological insights into moral theory, I saw that Kant could provide as rich and concrete a moral theory as anyone (p. ix).

Frierson's book is an attempt to reconcile the anthropology with Kant's transcendental freedom.

<sup>35</sup>So Wood (1999) sees Kant as anticipating the historical materialism of Marx (pp. 14, 225, 245). This is not an issue I will pursue in this paper, but for relevant criticism see Bielefeldt (2001, especially p. 450).

<sup>36</sup>This is the position taken in chapter 1 of Wood (1999). For further discussion see Baron (1995), who ultimately defends duty against the many charges against it.

moral demands away when they cease to be convenient (GW 4:390 ff.). It might be said, too, that the abstract form of the moral law is a consequence of the robustness Kant accords moral normativity, in making moral principles akin to principles in logic or mathematics. To conceive of moral demands as robust is a way of taking the domain of value very seriously, and that might be thought a virtue of the Kantian position. Again, that Kant's theory is on the face of it ahistorical might be important. Where more historically situated theories must answer worries about conservatism and relativism, Kant's theory seems fit to stand in critique of local practices and prejudices, including those to which Kant himself gives expression in the writings on history and anthropology.<sup>37</sup> Moreover, while it is surely true that Kant is not committed to an opposition between duty and inclination, the extent to which he can and should accommodate motives of sympathy and love should not be over-stated.<sup>38</sup> For Kant's characterisation of morality as categorical, as being such that moral demands must be heeded irrespective of inclination, including sympathetic inclination, is a distinctive claim that many moral theories will want to recover.<sup>39</sup> So while I recognise that there have been fruitful exchanges between, for example, Kantians and virtue ethicists, I would urge caution about accommodating the diverse theoretical commitments of other positions. It is important to appreciate that many of the most valuable insights of the

---

<sup>37</sup>It is common for Kantians to wield Kant's theory against his own unfortunate views about marriage, race or sexuality. See for example some of the moves in Section I of Korsgaard (1992).

<sup>38</sup>The claim that Kant is committed to a problematic opposition between duty and love is memorably encapsulated in Friedrich Schiller's satire of Kant on helping friends in need: "You must seek to despise them / And do with repugnance what duty bids you" (Schiller's satire is cited in Wood 1999, p. 28). A contemporary version of the objection may be found in Stocker (1976, p. 462). For an early defence of Kant against the objection, see Herman (1993b); for a pithy clarification of the issue, see Korsgaard (1997, p. xiii, fn. 6), and Loudon (1986, p. 487-8). I agree with Bielefeldt (2001, p. 449), that Wood overstates the reply to the objection when he claims that for Kant, "philanthropic love is *an indispensable ground of morality*" (1999, p. 39; italics in original).

<sup>39</sup>Scheffler (1994) has helpful discussion of Kant's challenge to what he calls motivational naturalism at pp. 61-72.

Kantian position are due to its distinctiveness.<sup>40</sup>

The structure of the paper is as follows. In Section 2.2, I locate metaphysics of morals within Kant's broader philosophical system. In Section 2.3, I lay out Kant's argument for undertaking metaphysics of morals. In Section 2.4, I clarify the positive role for empirical considerations in Kant's ethics. The first three sections form the basis for evaluating the interpretive trends at issue for the remainder of the paper. Section 2.5 considers the scope for incorporating virtue-ethical considerations in Kant's ethics, focussing on the work of Barbara Herman. I argue that from the point of view of Kant's systematic commitments, Herman's are possible Kantian moves, though they ought to be resisted on other Kantian grounds. Section 2.6 examines Allen Wood's attempt to read the anthropology into Kant's core theorising; while Section 2.7 examines Christine Korsgaard's attempt to read empirical considerations into Kant's account of the bindingness of morality. I argue that Wood's and Korsgaard's proposals, whatever their independent philosophical interest, are not plausibly Kantian proposals. I conclude, in Section 2.8, by showing why this is important to notice.

## 2.2 Metaphysics of morals

'Metaphysics' is Kant's term for a branch of philosophy.<sup>41</sup> It is a branch of pure, as opposed to empirical, philosophy, and it is the branch of pure philosophy concerned with synthetic a priori truths—substantive truths that are knowable on the basis of reason alone. The task of metaphysics is twofold. It is critically to investigate the capacity of reason to have substantive a priori knowledge, and it is to determine the body of substantive truths arising

---

<sup>40</sup>Sherman (1997, p. 2) makes a similar point, though it is an aside in a monograph that brings Kant into dialogue with Aristotle. Hursthouse (1999) voices a different view when she says of the various positions in normative moral philosophy: "Let us by all means stop caring about how we distinguish ourselves and welcome our agreements" (p. 7).

<sup>41</sup>His division of the areas of philosophy is prefigured in the first *Critique*, see A841/B869, and taken up in the preface to the *Groundwork*.



out of pure reason. Call these, for sake of clarity, metaphysics and metaphysics\*.<sup>42</sup> There is metaphysics of nature and metaphysics of morals, and in both cases, metaphysics must prepare the way for metaphysics\*.

The *Groundwork* is a work of metaphysics of morals. Its task is to prepare the ground, or lay the foundations, for the system of duties—the metaphysics of morals\*—promised to come later.<sup>43</sup> We are told that the task of moral foundations involves “nothing more than the identification and establishment of the supreme principle of morality” (GW 4:392).<sup>44</sup> ‘Identification’ and ‘establishment’ of the principle of morality are ways of describing the section divisions of the *Groundwork*. Sections I and II ‘identify’ the principle of morality, and its source, by analysing the deliverances of common human reason on the nature of morality and moral worth. From these deliverances, the Categorical Imperative is derived as the supreme principle and its ‘source’ is taken to lie in reason, and more particularly, in autonomy of the will. But it remains to be ‘established’ that there really is a principle of morality with the source and features analysed from common human reason. So Section III provides a ‘deduction’ of morality, and part of the burden of the deduction is to explain how morality can have its source in something radically independent of all human interests, namely, in autonomy of the will. What needs to be shown is that morality can be binding on agents though it does not rest on their antecedent desires or values. Both of these tasks,

---

<sup>42</sup>Kant does not always clearly distinguish between them. Guyer (1998a, p. xiv) notes that the phrase is used with a “dual sense”; and Gregor (1963, p. 4) recognises a more inclusive and a narrower use of the term. What I am calling ‘metaphysics’ “investigates the faculty of reason in regard to all pure *a priori* cognition”, while what I am calling ‘metaphysics\*’, lays out the “philosophical cognition of pure reason in systematic interconnection” (A841/B869). I take my gloss of metaphysics\* from the rendering of the translation in Beck (1960, p. 9). My concern in this paper is with metaphysics of morals, though I remark on metaphysics of morals\* in footnote 42.

<sup>43</sup>On the significance of the metaphors of ‘grounding’ and ‘foundations’ for the title, see Guyer (1998a, pp. xiii-iv). On this way of reading the title of the *Groundwork*, see Korsgaard (1997, p. x).

<sup>44</sup>I have adapted Gregor’s translation here in light of discussion with Jens Timmermann. ‘Identify’ is better than Gregor’s ‘search for’, because ‘Aufsuchung’ is a success term.

the analytic task of ‘identification’ and the synthetic task of ‘establishment’, are the work of moral foundations, or what comes to the same, metaphysics of morals. As I am using the terms ‘moral foundations’ and ‘metaphysics of morals’, then, they concern *both* the ground of morality and its bindingness.

How might we rephrase the concerns of metaphysics of morals in our terms? Metaphysics of morals is a form, as we might put it, of meta-ethical rationalism. And it is a form of meta-ethical rationalism in two ways, corresponding to the ‘identification’ and ‘establishment’ functions of the *Groundwork* described above. It is a form of rationalism i) because the ‘source’ of morality, that on the basis of which we have moral obligations, is reason. And it is a form of rationalism ii) inasmuch as it takes reason to be the explanation of the bindingness of the moral law—of its authority for us. In short, morality is binding on us not because we want to be moral, or because it serves our interests; we heed moral commands because they are demands of reason, and we recognise the authority of reason.

It is helpful to consider a contrast class—to consider the kind of theory Kant’s rationalism excludes. The stalking horse of the *Groundwork* is the kind of theory which takes the task of reason to figure out what human beings need in order to live well (GW 4:395). This form of Aristotelianism puts an “empirically conditioned” conception of reason at its basis (CPrR 5:16), empirically conditioned because reason is put in the service of the natural human desire for happiness or welfare. One of the tasks of critique in the practical philosophy is to prevent empirically conditioned reason from alone presuming to set the terms for what we should be doing. It is to prevent empirically conditioned reason from claiming “absolute rule”, and from “expressing itself in demands and commands that go quite beyond its sphere” (CPrR 5:16).<sup>45</sup> Empirically conditioned reason is the target of

---

<sup>45</sup>Kant gives a provisional argument against empirically conditioned reason to the effect that reason is inadequate to the task of securing happiness; lacking insight into our needs and what fulfills them, reason actually proliferates our needs and makes us ever more unsatisfied (GW 4:395). I take it that this is a relatively

practical critique just as speculative metaphysics is the target of theoretical critique; the difference is that reason shows opposite tendencies in the two cases. In the theoretical case, reason tends to stray beyond experience and claim knowledge of the super-sensible, while in the moral case, reason underestimates the practical reach of pure reason and tends towards empiricism. The reversal is not lost on Kant (CPrR 5:16).<sup>46</sup>

In sum, Kant's commitment to metaphysics of morals is a commitment to a substantive position in meta-ethics regarding the source and bindingness of morality. This commitment excludes theories which incorporate empirical considerations into the source and bindingness of morality, for example, theories which start with the good for human beings. I have not yet said *why* Kant is committed to metaphysics of morals, just *that* he is and what that commitment comes to. But as the discussion so far makes clear, the commitment is Kant's big move in moral philosophy. It is time now to say why he thinks it is needed.

### 2.3 Kant's Argument for Metaphysics of Morals

Why does Kant think that a metaphysics of morals is needed? Why must moral foundations proceed entirely a priori? Here is the characteristic answer Kant gives us:

[T]hat there must be such a philosophy [metaphysics of morals] is clear of itself from the common idea of duty and of moral laws. Everyone must grant that a law, if it is to hold morally, that is, as a ground of an obligation, must

---

weak argument against empirically conditioned reason on Kant's part, but that the case against it is not meant to rest on this argument alone. Kant gives a more forceful argument when he characterises morality as necessary and universal. We will come to this argument later.

<sup>46</sup>That pure reason has so positive a role in the practical philosophy may come as a surprise given that it is criticised so severely in the theoretical philosophy. Neiman (1994) thinks that to be surprised about this is to be confused about the nature of reason for Kant. In particular, it is to confuse the aim of reason with the aim of knowledge. Against this Neiman urges that, for Kant, reason is thoroughly practical. Taken this way, Neiman encourages us to think of the first *Critique* as correcting improper demands placed on reason—that it give theoretical answers to speculative questions—so that it may be given full sway, given primacy, in the practical domain.

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 (4:389).

The passage begins with the claim that we have an ordinary notion of moral law, and the suggestion is that because we think in terms of law, morality cannot have an empirical foundation—that, as the passage continues:

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.

The passage gives us the characteristics of moral laws that make for the exclusion of empirical considerations. Kant implies, since he says moral laws hold for all rational beings, that moral laws are universal, and Kant tells us outright that moral laws are necessary. What does necessity amount to in the present context? In general, for Kant, something is necessary if it could not be otherwise—if there are no possible worlds in which it is otherwise for beings with minds like ours.<sup>47</sup> While Kant presumably thinks that the principle of morality is the principle of morality in all conceivable worlds that are also moral worlds, the passage seems to give us a different, a further, way of thinking about moral necessity. Here Kant seems to treat necessity as a feature of the law’s grip on us; Kant says the law must be *heeded*. Moral requirements are practically necessary in the sense that they are rationally *binding* on agents.<sup>48</sup> Given Kant’s background assumptions about the connection between necessity, universality and the a priori, and between the a priori and reason, it follows

---

<sup>47</sup>That is, for beings whose intuition is sensible.

<sup>48</sup>Note that it is for beings like us, beings with a mixed rational-sensible nature, for whom the law has the character of bindingness. This is because, unlike angels and unlike God, subjective conditions may stand in the way of our acting in conformity with reason (GW 4:412). Kant says: “All imperatives are expressed by an *ought* and indicate by this the relation of an objective law of reason to a will that by its subjective constitution is not necessarily determined by it (a necessitation)” (GW 4:413) See also the preceding passage in the *Groundwork*.

that if morality is necessary and universal, it must have its source in reason.<sup>49</sup> So Kant's argument for excluding empirical considerations from moral foundations—for undertaking metaphysics—turns on the nature of morality: that morality is such that it is binding on agents (is practically necessary), and holds for all dependent rational beings (with universal scope).

## 2.4 The Positive Role for Empirical Considerations

We have been discussing metaphysics of morals as the preparatory or foundational part of moral philosophy. We have outlined the kinds of questions it enquires after, as well as the constraints that, as a division of pure philosophy, are imposed on answers to those questions. We have seen that it is Kant's conception of morality that drives the exclusion of empirical considerations from its foundations. The question we turn to now is: what is the positive role for empirical considerations in Kant's theory? We ask the question because we want an interpretive basis with which to adjudicate the claims of recent interpreters to the effect that Kant is not the aprioristic philosopher we have taken him to be.

Now it is uncontroversial that empirical considerations enter in 'anthropology'. According to the divisions of philosophy set out in the preface to the *Groundwork* (GW 4:388), anthropology is the empirical counterpart to metaphysics of morals. But, for Kant, anthropology is not strictly speaking a branch of moral philosophy at all. This is presumably because anthropology is not normative but, as we might put it, strategic and psychological. For example, a feasible task for anthropology is to institute what Kant calls an "aesthetic of morals": the vivid arousal of feelings of disgust and horror in the face of transgressions of

---

<sup>49</sup>Kant thinks that necessity and universality are hallmarks of the a priori. As Kant puts it in the first *Critique*, universality and necessity are "marks by means of which we can securely distinguish a pure cognition from an empirical one" (B3). Since Kant thinks we inquire a priori only into what reason itself contributes (Bxviii), that we have apriority shows that the source of that cognition is reason.

the moral law which serve to fortify commitment to duty (MM6:406).<sup>50</sup>

So far the picture is this: we have moral foundations which admit nothing empirical, and moral anthropology which depends on empirical facts about human beings, but is non-normative. Does anthropology exhaust the role of empirical considerations in Kant's theory? Indeed, can it? It would seem that more needs to be said. For while Kant insists that pure reason must determine the requirements of morality on its own, those requirements must be determinate, ultimately, for human beings. Consider that the principle of morality is to serve a deliberative function. On a standard reading, we are to formulate the maxim of a prospective course of action, and test whether the maxim can be universalised; if not, we cannot will to act.<sup>51</sup> Now in specifying our maxim (what we propose to do for some reason), we must specify the circumstances of our action and our motive. That specification would seem to involve empirical content. In this way, the application of the principle of morality seems to incorporate empirical considerations. Unlike anthropology, however, its results are normative: some such particular action is permissible, or must, or must not, be done. The application of the principle of morality would appear to lie somewhere between foundations and anthropology.<sup>52</sup>

So indeed, in *The Metaphysics of Morals*, Kant complicates the *Groundwork* divisions by envisaging a part of moral philosophy that is neither metaphysics nor anthropology, but includes empirical content to form an addendum to metaphysics. In fact, as I would put it,

---

<sup>50</sup>See also MM 6:217 where anthropology is said to concern "the development, spreading, and strengthening of moral principles (in education in schools and in popular instruction)".

<sup>51</sup>There is controversy over whether deliberation is indeed the purpose, or the sole purpose, of the Categorical Imperative. Plausibly, Kant does not intend us to test every maxim of action, though he does intend us to test some. This is the view taken by Loudon (1986), but for a different view, to which Loudon's paper is a reply, see O'Neill (1984).

<sup>52</sup>For the idea that it is in maxims of action that empirical (situational) content enters Kant's theory, see O'Neill (1985).

there seems to be several layers of addenda to metaphysics, moving from the more abstract to the more concrete, as more empirical content is incorporated. It is in the addendum to metaphysics that the application of the moral principle presumably belongs. And it is in the addendum to metaphysics that ethics falls into “casuistry” on matters such as being a responsible host, and being thrifty, on abstaining from flattery, and complaint. We are told that casuistry is not part of systematic philosophy, but is “woven into ethics in a fragmentary way” (MM 6:411). Kant marks this out, as it were, textually, by placing casuistical remarks in sections that are separate from the body of the text.<sup>53</sup> In this intermediate philosophical domain lies the task, also, of working out the duties that govern treatment of human beings of particular rank, age and sex, of particular states of learnedness, cultivation and health (MM 6:468-9). Again, Kant tells us that this application of the moral law “cannot be presented as sections of ethics and members of the *division* of a system (which must proceed *a priori* from a rational concept), but can only be appended to the system”. Kant adds that “[y]et even this application belongs to the complete presentation of the system” (MM 6:469). The qualification is presumably meant to signal that the duties to particular individuals, while they cannot be part of metaphysics in incorporating empirical content, still issue in necessary connections between a person’s will and the treatment of someone. They have to that extent the kind of normative content that makes them unsuitable for anthropology.<sup>54</sup>

What emerges is this. There is a sharp distinction between the foundations of morality, on the one hand, and the various dimensions of applying the principle of morality, on the other. The principle of morality is purely rational even if its application may draw on

---

<sup>53</sup>On the importance of observing the section divisions in *The Metaphysics of Morals*, the divisions between what is part of systematic philosophy and what is not, see Katja Maria Vogt (2008a, p. 236, p. 239 and ff.).

<sup>54</sup>On this point see Gregor (1963, p. 8).

experience.<sup>55</sup> What the principle commands on a given occasion is pure in the sense that the will is determined, not by inclination, but by the law. However, working out what one's duty involves applying the principle, and with that, empirical data (which for Kant is neither here nor there morally).<sup>56</sup> In sum, the role for empirical considerations in Kant's moral philosophy is as follows. Empirical considerations enter indisputably outside of metaphysics, for example, in the application of the moral law in concrete situations, and in anthropology. But empirical considerations in no way enter in the determination of the source or bindingness of morality.<sup>57</sup>

---

<sup>55</sup>It follows that it is not the consequences of a given action that rule out a maxim, but the contradiction that an agent would incur were he to act on it. I am grateful to Jens Timmermann for this point.

<sup>56</sup>I am grateful to Jens Timmermann for discussion on these points.

<sup>57</sup>There are further questions about the role of empirical considerations in what I have called metaphysics of morals\*: the system of duties arising out of pure reason. In the *Groundwork*, metaphysics of morals\* is said to be a branch of pure philosophy. But Kant made his systematic divisions of philosophy before he came to the task of actually setting out a system of duties. It is conceivable that in turning to the task, he came to see that more empirical content is needed for the specification of duties than he at first assumed. As he writes of the projected, comparable task in the theoretical philosophy, "An organon of pure reason would be a sum total of all those principles in accordance with which all pure *a priori* cognitions can be acquired and actually brought about. The exhaustive application of such an organon would create a system of pure reason. *But since that requires a lot, and it is still an open question whether such an amplification of our knowledge is possible at all and in what cases it would be possible*, we can regard a science of the mere estimation of pure reason, of its sources and boundaries, as the propaedeutic to the system of pure reason" (B25; long italics mine). It seems to me that Kant never wrote a metaphysics of morals\* as it was envisaged in the *Groundwork*, and that he was then unsure precisely what form it would take.

Insofar as we have what the *Groundwork* called a metaphysics of morals\*, it contains less an account of all the particular moral principles to which rational beings are subject, but, more abstractly, the systematic divisions into which those principles fall—divisions between duties to oneself and others, between perfect and imperfect duties, and between duties of 'right' and duties of 'virtue'. When Kant introduces the examples of duties in the *Groundwork*, he says they will be systematically set forth in a future metaphysics of morals\*. But it is worth noting that the future task is said to be not so much the elaboration of the examples, but the elaboration of the *divisions* to which those examples belong. Kant says, "It must be noted here that I reserve *the division of duties* [i.e. between perfect and imperfect] entirely for a future *Metaphysics of Morals*, so that the division here stands only as one adopted at my discretion (for the sake of arranging my examples)" (GW 4:421 fn; italics mine). This being so, we might urge that Kant never wrote a metaphysics of morals\* understood as a system of duties for rational beings. If the book with that title contains metaphysics\*, it contains it in virtue of the account of the systematic divisions among duties, and their associated notions. In this way, we may take *The Metaphysics of Morals* to be a strangely mixed kind of book. It is a mix of metaphysics as critique, of metaphysics\* in the sense just outlined, and of the various divisions of practical philosophy which lie *outside* of the system of metaphysics strictly speaking, and to which any mention of specifically *human* duties belongs.



## 2.5 Virtue-Ethical Lessons from *The Metaphysics of Morals*

Clearly, that empirical considerations enter in the extra-systematic reaches of moral philosophy bears on certain ways of thinking about Kant as an aprioristic moral philosopher, and some Kantians have fought against the charges of apriorism from precisely this point of view. Barbara Herman is a case in point, and I turn now to Herman as issuing the first of three interpretive challenges to Kant's apriorism.

Herman is interested in the *philosophy of* applied Kantian ethics, or in what Herman (1993a) calls "middle theory". Middle theory is so called because "it lies between the high theory of value and the low theory of applications"; middle theory "effects the translation of the basic conception of value in the principles of practical rationality into principles that fit the circumstances of human action, judgment and deliberation" (pp. 233, 236).<sup>58</sup> Since Kant's own work in this area is schematic, middle theory "provides the missing link in a reconstruction of Kantian ethics" (p. 233). It does so by supplying the necessary, supporting conditions for the Categorical Imperative to work as a test for maxims (1998, p. 143). Herman has sought to substantiate Kant's claim that the application of the Categorical Imperative requires judgment in light of experience (MM 6:411; GW 4:389). Her proposal is roughly as follows. In order to make use of the a priori principle of morality an agent must accurately formulate the maxim of his action. If he is not to conceive of his situation idiosyncratically, he must, among other things, have some understanding of the moral bearing of his situation. In Herman's terms, he must be able to single out what in his situation calls for "moral attention"—the "morally salient" marks and features of his action and circumstances. For Herman that presupposes a capacity for "moral perception" (what she also calls a "moral sensitivity"), as well as the possession of moral concepts, and these

---

<sup>58</sup>Page references are to the reprinted (collected) editions of Herman's papers.

will need to be acquired as part of a normal upbringing.<sup>59</sup>

Herman's work in middle theory in this way appeals to notions traditionally of interest to the virtue ethicist.<sup>60</sup> Middle theory, Herman says, makes room for a notion of character in Kant.<sup>61</sup> Let me note two qualifications about this move. Firstly, it is important to be clear that there are two ways a notion of character may be invoked. On the one hand, the idea might be that while what matters, morally speaking, is the exercise of practical reason, the exercise of practical reason presupposes the right formation of our desiderative natures.<sup>62</sup> On the other hand, the idea might be that what matters, morally speaking, is character, where character is understood as an assemblage of qualities or traits (the virtues of courage, temperance and so on), where these are defined, roughly, as separate dispositions to behave appropriately in relevant situations.<sup>63</sup> Whether Kant's theory can in principle accommodate a notion of character depends on whether it is invoked in the first or second way. Since the first gives no foundational role to character, it is at least consistent with Kant's systematic commitments. But the second can in no way feature in Kant's theory; in

---

<sup>59</sup>See Herman (1985, pp. 74-8), and Herman (1996, p. 1 and ff.).

<sup>60</sup>Or at least, to virtue ethicists working in a certain tradition. The tradition is Aristotelian, but in the interpretive vein of Wiggins (1980) and McDowell (1998). See the papers collected in the first part of McDowell (1998).

<sup>61</sup>Herman (1993b) tells us that: "Kant's notions of virtue and character are in no way peripheral to the understanding of moral judgment and action. We are able to consider the nature of a Kantian moral agent—what motives, feelings, thoughts, and commitments guide her deliberations and actions. There is then room to develop an account of moral personality that places moral activity within the ongoing practical commitments of a good life" (p. x). This position is developed in Herman (1983), but see also Herman (1996).

<sup>62</sup>See Bk. VI, Ch. 13 of the *Nicomachean Ethics* in Aristotle (1987). McDowell has emphasised the primacy of practical reason, and of virtue as necessary for practical reason, in contemporary discussions.

<sup>63</sup>The second is the more common conception of the role of character in contemporary virtue ethics. For influential formulations of the position along these lines, see Watson (1990) and Hursthouse (1999, ch. 1). Watson explicitly contrasts this approach with that of McDowell at (p. 466, fn. 7). These theories tend to make little mention of practical reason, or where practical reason is mentioned, it can be difficult to see how it is integrated into the overall picture. This may be because it is hard to see what (fundamental) role is left for character once reason enters the account. I am grateful to Katja Vogt for bringing me to see this point.

putting an empirical notion at the foundation of moral theory, it is ruled out, not least, by Kant's systematic commitments. ("Not least" because in failing to make moral action action from duty, it cannot be part of a Kantian proposal).<sup>64</sup>

The second qualification is that the notion of character, whether it is employed in the first or second way, is *not* Kant's own. To *have* character, or virtue, for Kant, is just to be the kind of person who commits himself to action from duty; it is to be someone who determines himself to act by thought of the moral law (MM 6:404).<sup>65</sup> In fact, the idea of virtue in *The Metaphysics of Morals* is explicitly contrasted with the more Aristotelian idea of virtue as "an *aptitude* and [...] long-standing *habit* of morally good actions acquired by practice" (MM 6:383).<sup>66</sup> To the extent Kant is interested in character, he is not interested in Aristotelian character.<sup>67</sup>

There are philosophical reasons for this. Firstly, for the Aristotelian style virtue ethicist, a person must have a good upbringing to get it right ethically speaking. But it is important for Kant that this not be so. It is part of Kant's Enlightenment appeal that no matter what one's training and background, everyone is capable of knowing what their duty is, and of doing

---

<sup>64</sup>It is sometimes complained that Kant's ethics suffers for lack of a role for character of the second kind. See, for example, the references to MacIntyre and Foot in Louden (1986, pp. 473-4). But we might side with O'Neill (1984) in thinking that the point should rather go the other way; that virtue theories which give a role to character of the second kind need to be corrected by the Kantian (and indeed the Aristotelian) emphasis on practical reason.

<sup>65</sup>As Kant puts it, virtue is "the moral strength of a human being's will in fulfilling his *duty*, a moral *constraint* through his own lawgiving reason, insofar as this constitutes itself an authority *executing* the law" (MM 6:405).

<sup>66</sup>A similar contrast is drawn in *Anthropology from a Pragmatic Point of View*. In one sense, character is an assemblage of particular traits, and in another character is the mark of a principled person, a person who acts from duty. Of the distinction, Kant tells us that "the first is the distinguishing mark of the human being as a sensible or natural being; the second is the distinguishing mark of the human being as a rational being endowed with freedom. The man of principles, from whom one knows what to expect, not from his instinct, for example, but from his will, has a character" (A 285).

<sup>67</sup>The distinctiveness of Kant's conception of virtue and character is for the most part well taken in the literature. It is remarked in many of the essays in Betzler (2008). See also Baxley (2010).

their duty. Secondly, where Herman suggests that our formation of practical reason requires the right formation of our desiderative natures, Kant would more likely see matters the other way. The thought would be that we must first settle on the right rational scheme—on the right principle—and cultivate our natures in terms of it. So much is suggested by his remarks about the educative function of anthropology (quoted above).<sup>68</sup> Thirdly, where Herman suggests that the proper formulation of maxims will require moral thinking that is responsive to local values and practices, she must contend with Kant’s notion of common human reason. For Kant, the Categorical Imperative is a formalisation of the implicit ideas that reason simply has, and it is not clear that this is meant to be socially articulated.

In sum, I would argue that Herman’s work in “middle theory”, while it invokes notions that require some care, amends Kant’s view in ways that are at least consistent with his system. In this Herman may be taken to draw out an important lesson, namely, that Kant permits empirical considerations to enter in the application of the moral law, that is, outside of his system. Still, while the rapprochement between Kantianism and Aristotelianism has the potential to address longstanding objections to the Kantian view, we should resist finding too easy a confluence of ideas.

## 2.6 A Foundational Role for Human Nature?

In the next two sections I turn to proposals which make what I take to be more controversial amendments. They are controversial in allowing empirical considerations to enter the account of the source of morality, on the one hand, and the bindingness of morality, on the other. It is to the first of these that I now turn.

In an influential paper, Allen Wood (1991, pp. 326-7) claims that ethics is concerned

---

<sup>68</sup>I am grateful to Wolfgang Mann for discussion on this point.

with the conduct of human beings, and accordingly, must be “based” on knowledge of human nature. Wood asks whether Kant’s ethics presents an exception to ethics so defined. It would appear to present an exception, indeed an objection, Wood says, given Kant’s commitment to metaphysics of morals. Against appearances, Wood argues that Kant’s ethics is “founded” on an empirical account of human nature, and, as such, presents no exception to his definition of ethics. “Even in a theory whose fundamental principle is a priori”, Wood tells us, “that account will be based on an empirical account of human nature, on an anthropology” (p. 326; see also p. 337). Wood takes this finding to license some radical interpretive conclusions: “To see Kantian ethics my way is to focus on very different issues from those which have usually occupied Kantian moral philosophers” (p. 345). He goes on:

My aim here is not to decide the issues I have just been raising, but only to legitimize them. They are issues about human nature and its historical destiny, falling entirely outside the scope of a Kantian ‘metaphysics of morals’, belonging instead to a critical examination of the anthropological foundations of Kantian moral theory. We cannot begin to evaluate these foundations, however, until we have first admitted their existence, and that requires us to overcome some traditional ways of looking at Kantian ethics (p. 347).

The claims are repeated in Wood (1999), where the interpretation is worked into a book length study.<sup>69</sup> Since the claims are radical, and roundly influential, we must look into Wood’s arguments.

Wood argues as follows. If we attend carefully to Kant’s argument for metaphysics of morals, we will see that it is a practical argument, and rests on empirical claims about the nature of human beings. Kant argues, for example, that human beings are beset by many

---

<sup>69</sup>“Kant’s position is grounded on a distinctive theory of human nature and history” (p. xiii); “Kant never meant to deny the essential place in ethics of an empirical study of human nature” (p. 10). Despite the continued presence of these claims, the monograph seems to be more circumspect than the early paper. For one, in accordance with the systematic divisions outlined above, the book is divided into two parts, *Metaphysical Foundations* and *Anthropological Applications*.

unruly inclinations, and that only the thought of pure reason, unmixed with empirical incentives and feelings, has an influence on the human heart sufficient to master them. By contrast, a theory which grounds morality in a mix of inclinations and feelings as well as rational concepts causes the human mind to waver about what is required of it (GW 4:410-11). Human beings are also incapable of following examples of virtuous conduct, so they need an a priori principle of morality to live by (GW 4:406). Moreover, the motive to morality must be the a priori motive of duty because only then can good actions be *reliably* produced (GW 4:390; 4:411). The purity of the moral law (and motive) serves these practical functions, and since the practical functions depend on empirical claims about human beings, Wood concludes: i) that anthropology drives the need for purity in moral foundations, so that ii) the basis of morality turns out to be human nature, after all (Wood, 1991, pp. 326-7).

What are we to make of i) and ii)? In a clear sense, even if i) is true, ii) in no way follows. That is, even if we need metaphysics of morals for psychological-cum-motivational reasons, it does not follow that metaphysics of morals is itself anthropological. The fact that we have an instrumental justification to pursue the pure justification of morality does not make the pure justification of morality itself instrumental. So it seems that Wood must be using the notions of “foundation” and “basis” here in a rather special sense. Indeed, Wood tells us that he means “foundations” in “a somewhat larger sense” than is usually intended, so that in speaking of the foundations of morality he is speaking of the significance of morality and its role in human life (p. 326). Now of course philosophers can define terms as they see fit, but since the notions of “foundation”, “grounding” and “basis” are technical terms in Kantian ethics, doing so here is at best highly misleading—all the more so given the dialectical set up of Wood’s paper (see above).

Moreover, i) needs to be qualified. Wood is quite right that Kant thinks there are

practical reasons for undertaking metaphysics of morals, and right that these rest on claims about the nature of human beings. But the argument for metaphysics of morals *cannot* rest on the practical argument, and it is not meant to. It *cannot* rest on the practical argument because the practical argument depends on contingent facts about human beings, facts that may have been otherwise. Human beings may have been constituted so that we had no psychological need for the principle of morality to be formulated a priori, but the principle of morality would still have been a priori for that. Further, the argument for metaphysics of morals is not meant to rest on the practical argument, but on an argument to which Wood here makes no mention, viz., the argument from the character of the moral law.<sup>70</sup> In confining his attention to Kant's practical reasons for undertaking metaphysics of morals, Wood leaves out of the account what I take to be the real philosophical impetus behind Kant's project. As I argued above, Kant is not just concerned to undertake metaphysics of morals for anthropological reasons, Kant is concerned to undertake metaphysics of morals because of what he thinks morality *is*.

## 2.7 Bindingness Without Metaphysics?

I turn now to a final attempt to incorporate empirical content into Kant's account of the foundations of morality, this time into Kant's account of bindingness. While there have been a number of efforts to do this on the part of both neo-Kantians and Kant interpreters, I will be concerned with the efforts, in particular, of Christine Korsgaard. Korsgaard's view has been articulated in a number of important publications over many years.<sup>71</sup> It is not my

---

<sup>70</sup>In the monograph, in the course of laying out Kant's foundations properly so called, Wood (1999, ch. 2) describes the argument from the character of morality. That argument is not brought into contact with the earlier account of Kant's practical argument, and it is unclear how Wood now understands their relationship.

<sup>71</sup>See the essays collected in Korsgaard (1996a), the Tanner lectures published as Korsgaard (1996d), and more recently the Locke lectures published as Korsgaard (2009).

concern here to give a detailed account of Korsgaard's view, nor to trace out the relationship between the different versions.<sup>72</sup> Rather, I want to examine the first big move in Korsgaard's response to what she calls "the normative question", a response Korsgaard marshals from lines of thought in Kant; and I want to show how Korsgaard's response fails as a Kantian proposal in violating Kant's basic commitment to metaphysics of morals.

The normative question is the question of how moral requirements are binding on us, where that may be understood as the question of how moral requirements are both authoritative and motivating independently of what we may happen to want. This is the question Kant asks in the third part of the *Groundwork*, and in many ways Korsgaard's response is an extended reconstruction of Kant's first moves. Kant argues (i) that in willing to act we cannot but take ourselves to be free agents; and that (ii) the notion of free agency in question is that of giving the law to ourselves, so that freedom really is autonomy. Kant has already argued for an equivalence between autonomy and morality.<sup>73</sup> So from a would-be datum about our agency—that we must take ourselves to be free if we are to act all—morality is taken to follow.<sup>74</sup>

For Korsgaard, the normative question makes itself felt from the first person point of view of deliberating over what to do, and in particular, deliberating over whether to satisfy the demands of morality when doing so would come at some cost to the agent. What is wanted is an explanation of our reasons to act in accordance with morality which makes those reasons salient from the agent's point of view. It must be clear how the demands of

---

<sup>72</sup>This has been more than ably carried out elsewhere. See especially Fitzpatrick (2005), but also O'Hagan (2004).

<sup>73</sup>This is not meant to be a close textual analysis of the opening two sections of *Groundwork* III (GW 4:446-4:449). My sketch follows Korsgaard's own account of Kant's argument in Korsgaard (1996c).

<sup>74</sup>Or at least, morality is taken to follow "analytically". Kant thinks it still needs to be proven "synthetically", and that is what the remainder of the third part of the *Groundwork* is taken up with. This is unimportant for present purposes.



morality have normative force *for the agent*. Korsgaard begins with our agency because she takes a concern with agency to be given for human beings; our agency is something for which we cannot reasonably ask the question, But why should I be an agent? Since we can guarantee the normative force of our agency, to the extent moral principles can be shown to be a function of agency, we will guarantee the normative force of moral principles.<sup>75</sup>

In what sense is agency a given concern for human beings? Sometimes Korsgaard writes as if agency is a given concern because agency is inescapable or “practically necessary” for us. That is to return to the Kantian thought that we cannot act except under the idea of freedom. As Korsgaard (2009) puts it, “Human beings are *condemned* to choice and action. Maybe you think you can avoid it, by resolutely standing still, refusing to act, refusing to move. But it’s no use, for that will be something you have chosen to do, and then you will have acted after all. Choosing not to act makes not acting a kind of action, makes it something that you do” (I.I.I).<sup>76</sup> The problem with this response is that the fact that something is inescapable for us does not guarantee that it has normative significance for us. We may be forced to be agents and yet regard our agency rather dimly, so that our sense of normative significance is quite disconnected from our situation as actors in the world.<sup>77</sup> To answer the normative question, Korsgaard needs agency to be an inescapable *normative* concern for us.<sup>78</sup>

Given this problem, a more promising response may be found in Korsgaard’s remarks

---

<sup>75</sup>See Fitzpatrick (2005, section I).

<sup>76</sup>This idea is given extended treatment in Korsgaard (1996c).

<sup>77</sup>Enoch (2006) develops this criticism of Korsgaard at some length. As he puts it, Korsgaard’s interlocutor may say: “Perhaps [...] I cannot opt out of the game of agency, but I can certainly play it half-heartedly, indeed under protest, without accepting the aims purportedly constitutive of it as mine” (p. 188).

<sup>78</sup>See Enoch (2006): “The kind of necessity the game of agency has to enjoy in order to solve the problem we are now in is *normative* necessity” (p. 188).

about practical identity and identification.<sup>79</sup> Here the thought is not so much that we are constitutively bound to be agents, but that we are constitutively bound to take a normative interest in our agency. We must not only act, but see our actions as extensions of ourselves, and not, moreover, just any old self, but ourselves considered under the auspices of a practical identity that we care about. As Korsgaard takes over themes from Harry Frankfurt (1971), we must act on reasons with which we identify.<sup>80</sup> And now the thought is that since we care about our reasons, care about being agents, if it can be shown that this care commits us to acting morally, then it can be shown that we really are bound to act as morality requires. Taken this way, Korsgaard's solution to the normative problem is to posit an essential connection between morality and caring, not just any care, but a care that is constitutive of our humanity.

As will be clear, Korsgaard means her proposal to be an extension of Kant's own solution to the normative problem. There are ways of reading part III of the *Groundwork* which have Korsgaard's solution coming close to Kant's own. For it can look as if Kant is saying something like the following there. We are bound by morality because we must take ourselves to be free, where in doing so we must take ourselves to be members of a noumenal world whose laws are the laws of morality. Since we are mixed beings—since we also have a phenomenal nature—the question is how the laws of the noumenal world are authoritative for the phenomenal world. And then the thought might seem to be that the laws of the noumenal world are regarded as authoritative because as phenomenal beings we *aspire* to being noumenal beings. This reading is encouraged by Kant's confirmation of the deduction of morality in common human reason (GW 4:454). Kant pictures “even the most hardened scoundrel” as aspiring to be the better person he knows himself capable of being:

---

<sup>79</sup>See Lecture 3 of Korsgaard (1996d).

<sup>80</sup>See Korsgaard (1996d, p. 99, fn. 8).

This better person, however, he believes himself to be when he transfers himself to the standpoint of a member of the world of understanding, as the idea of freedom, that is, of independence from *determining* causes of the world of sense, constrains him involuntarily to do; and from that standpoint he is conscious of a good will that, by his own acknowledgements constitutes the law for his evil will as a member of the world of sense—a law of whose authority he is cognizant even while he transgresses it (GW 4:454-5).

Since Kant's deduction in *Groundwork* III is notoriously difficult, it is very natural to turn to this more commonsensical formulation, so that Kant's solution to the normative problem turns out to be that morality is binding on us because we are motivated by a self-ideal.<sup>81</sup> That is, morality gets its grip on us in virtue of the desire of our phenomenal self to live up to our better self—to our noumenal nature.

It should be clear from the foregoing, however, that this cannot be Kant's view. It cannot be Kant's view because it rests the explanation of morality on something empirical, namely, the desires of our phenomenal self. For the same reason, whatever the independent merits of Korsgaard's solution to the normative problem, it cannot be Kant's solution. This is because, while it may secure the bindingness of morality in some sense, it does so by appealing to caring, where care, following Frankfurt, is explained in the naturalistically respectable terms of belief and desire.<sup>82</sup>

In the end Kant may not give us a satisfying explanation of categorical bindingness. In the *Groundwork* he takes an explanation of the interest we take in morality to be beyond the limits of moral inquiry (4:461-2); and in the second *Critique* he faces up to the limits of explanation by telling us that categorical bindingness is simply *a fact of reason*.<sup>83</sup> It is just

---

<sup>81</sup>I take this to be the reading of Kant favoured by Velleman (2006).

<sup>82</sup>For a criticism of Korsgaard on grounds that she gives, like Frankfurt, a metaphysically deflationary account of Kant's notion of the will, see Herman (2002, especially section III).

<sup>83</sup>But see, for example, O'Neill's (1989) suggestive postscript on pp. 64-5. See also chapter 5 of Franks (2005).

a given that we take ourselves to be bound by morality. However unsatisfying, unlike the route taken by Korsgaard in positing categorical cares, and unlike the route which explains bindingness in terms of motivation by a self-ideal, to appeal to a fact of reason at least has the advantage of being consistent with Kant's commitment to metaphysics of morals.

## 2.8 Conclusion

Let me close by saying that my criticism of Wood's and Korsgaard's proposals as not tenable *Kantian* proposals is no mere matter of terminology. It has been the concern of this paper to argue that Kant's commitment to metaphysics of morals is basic to the kind of project he is engaged in in moral philosophy. His commitment to metaphysics of morals follows from the kind of thing he takes morality to be. Now Kant is at his most fervent in those parts of his writing where he insists on a methodological advance over his predecessors—over those who, despite their quibbles with one another, are united by a shared oversight or blunder. In the critical philosophy, that advance consists in limiting the aspirations of pure reason to know by examining the preconditions of knowledge itself. The result is a thorough dressing down of traditional metaphysics—that great site of “mock conduct” and “groping among mere concepts” (Bxv)—and, against empiricism, the provision of a role for reason in working out the necessary conditions for experience. In the practical philosophy Kant also dismisses the fantastically ambitious optimism of some forms of rationalism. This shows itself in the fact that morality is in the end an idea of reason, and resistant to proof. But Kant's main target here is not reason's propensity to *transcend* experience, but as we might put it, its *failure* to do so. Kant's particular target are moral theories which, in blending rational and empirical considerations, “substitute for morality a bastard patched up from limbs of quite diverse ancestry” (GW 4:426). The result is a thorough rebuke to the moral credentials of empirically conditioned practical reason, and a steadfast commitment to the

rational purity of moral foundations. In incorporating empirical considerations into the foundations of Kantian ethics, Korsgaard and Wood violate the very move which Kant takes to be his singular contribution to moral philosophy.<sup>84</sup>

---

<sup>84</sup>This paper grew out of Katja Vogt's graduate seminar on Kant and his critics in the Spring of 2008. I am hugely indebted to Katja for emphasising Kant's commitment to metaphysics of morals in that seminar, and for many subsequent conversations. I owe a large debt to Wolfgang Mann, who provided incisive editorial suggestions on an earlier draft, and helpful discussion on several points. I am grateful also to Jens Timmermann, David Velleman, Joseph Raz, Katie Gasdaglis, and Brad Weslake.

### 3 Responsibility and the Value of Intelligible Beings

#### 3.1 Introduction

I begin with Kant's distinction between two standpoints we can take on human beings. We can regard them practically, as agents, or we can regard them explanatorily, as things in nature. Taken in the first way, human beings are sources of thought and action; thoughts are in some sense *theirs* to think and direct, conduct is theirs to control. Taken in the second way, trains of thought and courses of conduct are set by the promptings of nature; it is not *they* but *it* that makes things happen. Call the first the *intelligible* standpoint and the second the *sensible* standpoint.<sup>85</sup>

In the third part of the *Groundwork of the Metaphysics of Morals*, the intelligible standpoint emerges in the first place first-personally. It emerges to accommodate a would-be practical datum. The datum is that we cannot act except *under the idea of freedom* (4:448). That means that when we are deliberating about what to do we must regard our deliberation, our reason, as authoring the terms of our conduct. For Kant that is to say that we must, as a matter of practical necessity, regard our wills as free. In so regarding our wills we are constrained to take the intelligible standpoint on ourselves, to think of ourselves as members of an intelligible world. Kant takes this perspective to exclude a perspective on ourselves as sensible or phenomenal beings. For he thinks we cannot both reason about what to do and take our reason, as he puts it, to “receive direction from any other quarter”, in particular not the quarter of our sensible nature (4:448). We must take our reason to author the terms

---

<sup>85</sup>Kant talks of the two standpoints in the *Groundwork of the Metaphysics of Morals* (4:452): A rational being is said to have “two standpoints from which he can regard himself [...]”. The talk of standpoints here should not be confused with the deflationary two-aspect interpretation of Kant's idealism according to which noumena and phenomena are two aspects of the same thing. That is, talk of two standpoints should not be taken to be metaphysically innocuous, and in my account of Kant's view I will move back and forth between talk of the standpoint on another as an intelligible being and talk of his *being* an intelligible being. Jens Timmerman (2007) is helpful on this point (see p. 133, fn. 30).

of our conduct independently from sensible influence. The exclusion between the two standpoints is taken to hold in the other direction, too: from a standpoint on ourselves as sensible beings our actions are necessarily determined by natural causal laws—our agency does not here come into view.

### 3.1.1 Regarding Others as Intelligible Beings

For the Kant of the *Groundwork*, then, it is primarily ourselves whom we must regard practically as intelligible beings, for it is by the necessity of acting under the idea of freedom, if we are to act at all, that we are brought to the intelligible standpoint. But what now about our ways of regarding other people? Does Kant think we are to assume the practical standpoint on them—to regard *other people* as intelligible beings? And would so regarding them exclude a standpoint on them as sensible beings? That is, would considerations about another’s sensible nature be excluded from our practical ways of relating to, and of our deliberations concerning, them? As the two standpoints are distinguished in *Groundwork III* it is not clear. It is clear that all rational beings must regard *themselves* as intelligible beings from the standpoint of action, but that is not the same as saying that *we* must so regard them, and exclusively for practical purposes. At least, Kant does not here put things this way.

Still, we can see how he may be so interpreted, for we might read such a prescription back into the second part of the *Groundwork* where we encounter the idea that human beings have value of a singular kind, and that morality is a matter of responding in the right way to that singular value. Kant tells us that the ground of the value of humanity is rational nature and that we must in all of our actions relate to human beings as rational beings, that doing so is to respect them (4:428). What it is to relate to human beings as rational beings will depend on how we understand rational nature, in turn. But one constraint on any

Kantian account of the value of rational nature is that it cannot be an empirical property. So much is guaranteed by the fact that Kant takes rational nature to be shared by human beings, angels and rational aliens—by the class of all conceivable rational beings. We can put this by saying that rational nature is not sensible but intelligible for Kant. Human beings have the singular value they have in virtue of their intelligible nature. So if we are required to appreciate and respond to the value of human beings, then we must appreciate and respond to them as intelligible beings. Doing so will be part of what it is to value them in the right way.<sup>86</sup>

### 3.1.2 A Dilemma About Responsibility

This is one route to the thought that the Kantian theory of value entails that we value others as intelligible beings. Depending on our conception of the ground of value (depending on our conception of rational nature), there are different ways of spelling out what that comes to. On one such spelling out valuing persons as intelligible beings involves appreciating them as sources of thought and action—as free agents. This is the kind of proposal with which I will be concerned in what follows. In particular, I will be concerned with two contemporary Kantian accounts on the part of David Velleman (1999) and Christine Korsgaard (1992) to the effect that we must always value others as intelligible bearers of freedom. I will argue that notwithstanding their evident attractiveness, the views have striking consequences for our interpersonal relations, in particular, for the part of our interpersonal relations in which we hold people responsible for their actions. I will raise a dilemma to bring this out. I will argue that Velleman's view invites the consequence that we must *never* hold people

---

<sup>86</sup>That we are to regard others as intelligible beings is perhaps most explicit in the Introduction to the Paralogisms in the *Critique of Pure Reason* where Kant conceives of our relation to others (to other *persons*) always and automatically as agents. Likewise, in the Third Antinomy of the *Critique of Practical Reason* Kant appears to give a completely general account of freedom so that its ascription is not in the first place first-personal. There we need not, as in the *Groundwork*, infer his position about the third-person case.



responsible for wrongdoing, while Korsgaard's view entails that we must *always* hold people responsible for wrongdoing. While they each give us positive reasons to accept one or other of these consequences, I argue that neither should be accepted for neither capture important moral aspects of relating to others as sensible beings.

### 3.1.3 Kant's Dilemma

It will be helpful to trace the lines of thought in Kant which lead to the two horns of the dilemma. For it is one to which Kant may be thought committed on account of his views of freedom and agency. The views of Velleman and Korsgaard are vulnerable to the dilemma inasmuch as they inherit (versions of) these lines of thought from Kant.

I begin with the first horn. It arises in Kant's attempt to prove that a morality with the characteristics he explicates indeed applies to us—is valid for us. Kant's first, and for present purposes crucial, move in this attempt is to argue that if we presuppose freedom then the validity of morality follows analytically. It follows analytically because freedom and morality are "reciprocal concepts": "a free will and a will under moral laws are one and the same" (4:447).<sup>87</sup> Not many people are convinced by Kant's argument to this conclusion but, as we will see later, it is a crucial part of the deduction of morality in *Groundwork III*.<sup>88</sup> It yields a conception of freedom, positive freedom, as acting in accordance with moral laws. If we add the premise that freedom is a necessary condition on responsibility, as Kant thinks it

---

<sup>87</sup>The analytic connection between morality and freedom is in outline the following: rational beings make things happen by willing them, and for the will to set the terms of action it must be free from constraint, negatively free—it cannot be acted upon by physical forces, it cannot be subjected by inclination—the will must cause itself. Since the concept of a cause entails that of laws, for Kant, and since the will is free it cannot be a natural law—it must be that the will gives the law to itself. The will is autonomous and Kant has already argued that autonomy is a formulation of the moral law (4:446-7). So Kant tells us, "a free will and a will under moral laws are one and the same" (4:446-7).

<sup>88</sup>The argument is rehearsed in the *Critique of Practical Reason* at 5:33.

is,<sup>89</sup> then we have the conclusion that a person is never responsible for wrongdoing. This is the first horn of the dilemma.

The second horn emerges from a natural move in response to the first. The move is to say that freedom is not acting in accordance with moral laws, freedom is being *capable* of so acting.<sup>90</sup> If a will is free that *can* act as the moral law demands, and if freedom is required for responsibility, then we are responsible for immoral actions insofar as we could have acted morally. If we add the premise that we are *always* capable of acting as morality requires, then it follows that we are *always* responsible for wrongdoing. The additional premise is licensed by a central Kantian contention, and a core part of Kant's method, namely that everyone can work out what morality requires.<sup>91</sup> And meeting these requirements is no less possible: "[t]o satisfy the categorical command of morality is within everyone's power at all times" (5:37).<sup>92</sup> It follows that inasmuch as we have done something wrong, we are always

---

<sup>89</sup>At least, this is a way of rephrasing his position. Kant does not use 'responsibility' (Verantwortung) as a technical term in the *Groundwork*. Rather the thought is that we are free when we are the source of our actions, when we self-legislate, and so initiate a causal train. Actions that we initiate in this way are then properly speaking attributable to us.

<sup>90</sup>This account of freedom is implicit in Kant's discussion in the 'Critical Elucidation of the Analytic of Pure Practical Reason' in the *Critique of Practical Reason*.

<sup>91</sup>Kant gives a pithy statement of this commitment in the *Critique of Practical Reason*: "the moral law commands compliance from everyone, and indeed the most exact compliance. Appraising what is to be done in accordance with it must, therefore, not be so difficult that the most common and unpracticed understanding should not know how to go about it, even without worldly prudence" (5:26).

<sup>92</sup>Likewise, "to command morality under the name of duty is quite reasonable...for in regard to this, what [a person] wills to do, that he also can do" (5:37).

responsible;<sup>93</sup> there are no excuses, no taking account of agential frailty.<sup>94</sup>

So much for the dilemma. I do not take either side to be obviously misbegotten. Indeed, I will now try to motivate first the one side and then the other. I will do so by considering views I take to be independently interesting and attractive. These are the views of Velleman and Korsgaard. They by no means exhaust the Kantian positions in the literature and as we will see they in some ways depart from and extend Kant's views. I choose them not because I take them to be faithful in all respects to Kant, but because they give forceful expression to central features of his thought. But I will argue that neither view is in the end acceptable for neither take moral account of important dimensions of our sensible nature. That in any event will be my diagnosis, and I will take it to countenance a disjunctive conclusion: either we uphold a dual conception of our nature as intelligible and sensible beings but

---

<sup>93</sup>There is a question about what Kant would take the scope of responsibility for wrongdoing to be. When he tells us that each of us is capable of doing what morality requires, does he mean that we are all capable of doing what is in accordance with morality, or more strongly, that we are all capable of doing what is in accordance with morality *from the motive of duty*? It might seem obvious, given that only actions from duty have moral worth, that Kant is committed to the second reading. But if Kant thinks we cannot be sure what precisely motivates an action, not even in our own case, then it might seem he holds the first sort of position: within the scope of praise and blame is only whether someone has acted in accordance with the moral law. This would take the duty of perfecting our own motives outside the scope of responsibility, which is perhaps quite right.

<sup>94</sup>Kant explicitly addresses the problem of responsibility for wrongdoing in the *Critique of Practical Reason*. He raises it as part of the problem of free will and determinism:

If I say of a human being who commits a theft that this deed is, in accordance with the natural law of causality, a necessary result of determining grounds in preceding time, then it was impossible that it could have been left undone; how then can appraisal in accordance with the moral law make any change in it and suppose that it could have been omitted because the law says that it ought to have been omitted?

He answers that it can be said to have been omitted because a person regards himself as “determinable only through laws that he gives to himself by reason”, so that every action is “to be regarded in the consciousness of his intelligible existence as nothing but the consequence...of his causality as *noumenon*” (5:97).

So considered, a rational being can now rightly say of every unlawful action he performed that he could have omitted it... (5:98)

Moral appraisal in this way sets its store by the potential of everyone to realise their freedom and be moral.

uphold it so that the relevant parts of sensible nature are not excluded from the valuation of intelligible nature; or we reject the dual conception of our nature. I will close with reasons to prefer the second conclusion to the first.

### 3.2 Velleman on Valuing Others as Intelligible Beings

I begin with David Velleman's in many ways deeply Kantian account of valuing persons in 'Love as a Moral Emotion'. Velleman is centrally concerned with the evaluative attitudes we are to take towards *persons*. My question is what is implied by his account for the evaluative attitudes we are to take towards the *actions* of persons, where that includes evaluating their actions as good and bad, right and wrong, and making ascriptions of responsibility for wrongdoing. I will begin by foregrounding Velleman's account of valuing persons, and then I will draw out an implication for responsibility for wrongdoing.

#### 3.2.1 The Pre-existing Value of Intelligible Beings

I take the heart of Velleman's conception of valuing persons to be that we must fully wrap our minds around, fully understand, the singular value that they have. The idea that persons have value is given in the *Groundwork* as the idea that persons are ends-in-themselves. Persons are said to have a nature, rational nature, that marks them out as ends-in-themselves, and makes them singularly ("absolutely") valuable (4:428-9). That it is the nature of persons that gives them value means that persons have value independently of any and all empirical properties, and independently of moral credentials. So persons, as Velleman puts it, are "objects of motivating attitudes that regard and value them as they already are" (p. 357).<sup>95</sup>

---

<sup>95</sup>We are familiar with the passages in Kant where he takes even those who behave terribly, even the vicious, to be objects of respect. (See for example (6:463) of the *Metaphysics of Morals*). Plausibly, it is this aspect of the ends-in-themselves idea which underwrites the classical Kantian view. It says that respect is owed to all persons, not in virtue of merit but because persons have a nature that gives them value. Recently, commentators have

We saw above that rational nature for Kant is not an empirical but an intelligible property, and Velleman takes the object of value to be precisely “the intelligible essence of a person” (p. 344). In this he takes over Kant’s description of intelligible nature as a person’s true or proper self—what a person is in himself rather than what, as an empirical human subject, he merely appears to be (Velleman, p. 348; Kant 4:457-8). In so describing the object of value the terms are set for our task as valuers of persons. The kind of value that persons have warrants a mode of appreciation “in which we submit to the object’s reality” (p. 367), and that submission is such that we “stan[d] back in appreciation of the rational creature he is” (p. 358). Our task as valuers, at least in the first place, is to be right appreciators—to appreciate persons for the value they *really* have.

### 3.2.2 Right Appreciation: Respect and Love

For Velleman, right appreciation finds expression in the attitudes of both respect and love. Velleman’s paper is in part an account and an endorsement of the theory of value that Kant developed for respect, but Velleman uses the account of respect to propose a theory of value for love. His central thesis is that the object of respect and love both is the Kantian intelligible subject, so that respect and love are different modes of paying tribute to that subject. The relevant difference between the attitudes for present purposes is that respect involves a kind of intellectual grasp, an idea of, the intelligible nature of persons, while love involves appreciation of that intelligible nature as something manifest in the empirical persona. In this last Velleman departs in a significant way from Kant for whom it would not make sense to try to cognise the intelligible subject. Such an attempt would be for reason to

---

challenged the classical conception of Kant’s view. Stephen Darwall (2008) has drawn attention to passages where Kant seems to take more like the merit view, though Darwall does not in the end defend this view as Kant’s. The merit interpretation has been defended by Richard Dean (2006). One way to challenge the merit interpretation would be to underscore the theory of absolute value which grounds respect for Kant. I take this up in ‘Absolute Value and the Specialness of Human Beings’.

“impotently flap its wings without moving from the spot in the space, which is empty for it, of transcendent concepts called the intelligible world, and so lose itself among phantoms” (4:462). And yet I think Velleman’s adaptation lends the Kantian theory of valuation some intuitive traction. The idea is that while we may have the idea that all persons have a special kind of value, when we interact with them that value may fully come to strike us. We may be brought to a striking realisation of the humanity in them, striking because heretofore their humanity was a mere idea to us. We might think of this as a dawning of personhood. What occasions the full appreciation of personhood for Velleman is a person’s empirical behaviours and qualities, but importantly, these empirical properties merely symbolise or serve as conduits for the real source of value, namely, intelligible nature.

In this way Velleman’s theory of valuing persons gives some, albeit symbolic, accommodation to the sensible features of human beings. At least when we love them, we appreciate the intelligible nature of persons by attending to their sensible nature. But it seems that in this there is an asymmetry between the good and bad in sensible nature, between those aspects of the sensible being that reveal intelligible nature and those aspects that do not. For it seems Velleman is committed to excluding the bad aspects of sensible nature from proper valuation of persons.

### 3.2.3 Exclusion of the Bad in Sensible Nature

To see why, consider more precisely what Velleman takes the object of valuation to be. It is “the ideal of a rational will”; a will “that must ‘think itself into the intelligible world’”, as Velleman quotes Kant from the Groundwork, “as the bearer of freedom, which cannot be found in the sensible order” (p. 348). Velleman is identifying the object of valuation as the being Kant introduces in order to deduce morality. To identify the properties of that being

we need to say more about the deduction.<sup>96</sup> We have already encountered the first move of the deduction: it is the move which says that if we presuppose freedom—if we presuppose that we determine ourselves to activity independently of external sensible influence—then morality follows analytically.<sup>97</sup> But supposing morality may be derived from freedom, we still need to establish that the condition of morality can be satisfied: that rational beings, we human beings among them, can be free (4: 449). And what reason have we to believe that?

Kant gives a reason grounded in our practical lives, and it is here that we meet the would-be practical datum: that we can act only by taking ourselves to author the terms of our conduct independently of natural causal laws. Now Kant thinks we need to tell a story about how this practical datum is at least not dismissible out of hand, for it is threatened by our knowledge that all of our actions are necessarily determined. The story is that our deliberative perspective points to our membership in an intelligible world with its very own laws. So we have the picture of two realms: the natural realm governed by laws of nature, and the intelligible realm governed by laws of freedom. If we were purely intelligible beings then, paradoxically as it sounds, we would be determined to be free. And given the analytic identity between freedom and morality, that means that we would be

---

<sup>96</sup>Velleman comes to his account of the object of valuation by way of an interpretive puzzle. The puzzle is how persons can be the object of Kantian respect when Kant initially formulates the object as the moral law, and not persons (4:400). Velleman's solution is that respect for law is merely prefatory for what, more fully spelled out, is respect-for-persons-who-are-lawgivers. Respect for law is respect for persons because the law is the law of the rational will, and the will is ours. (See Jens Timmermann (2007) on 'Reverence', p. 182). A direct piece of textual support for Velleman's reading is at 4:440:

Our own will insofar as it would act only under the condition of a possible giving of universal law through its maxims—this will possible for us in idea—is the proper object of respect; and the dignity of humanity consists just in this capacity to give universal law though with the condition of being itself subject to this very lawgiving.

I am not here concerned with Velleman's interpretive puzzle, though I think his interpretation is the right one. In what immediately follows I will trace a short path through the *Groundwork* which bears out, while it does not argue for, Velleman's reading.

<sup>97</sup>See the argument in footnote 2 above.

determined to act for the sake of the moral law. On the other hand, if we were purely natural beings then we would act always for the sake of our desires and inclinations. But human beings are members of both worlds. The question is how the laws of the intelligible world are authoritative for us. Kant says rather elliptically that the laws of the intelligible world “ground” the laws of the phenomenal world (4:453). We can draw out the thought by appealing to the confirmation of the deduction in common human reason (4:454-5). We take the laws of our intelligible will to be authoritative for our sensible will: the intelligible will serves and is acknowledged by us to be an ideal for the sensible will. Morality is now said to be grounded in that ideal will.

The relevant upshot for our purposes is that the object of valuation on Velleman’s view, the ideal of a rational will, is the bearer of freedom and morality *both*. It is the subject which must think itself into an intelligible world with its very own non-natural (moral) laws. The object of valuation is in Kant’s sense positively free. That means that the non-ideal sensible will, the bearer of inclination and desire and the source of the bad in sensible nature, is unfree. It is for that reason not properly put to the *person’s* account.

We have in this way the following grounds for excluding the bad in sensible nature from proper valuation of persons. When we respect someone, as is requisite, we are to direct our thought and attention to their intelligible natures. Intelligible subjects are the bearers of freedom and morality both, so we are to direct our thought and attention to what we might call the core of goodness in them. All empirical considerations are irrelevant to the mode of valuation proper to respect, including the bad aspects of sensible nature. When we love someone, on the other hand, an optional mode of valuation, we are to direct our thought and attention to intelligible nature, but we are to exclude only the bad features of sensible nature. Since the bad in sensible nature is not caused by the intelligible subject, the source of morality and freedom, it is not part of the empirical persona through which the



intelligible self is revealed. The bad in sensible nature is for that reason irrelevant to how we see persons.

### 3.2.4 Love as an Ideal for Interpersonal Relations

I have just said that love is an optional mode of valuation for Velleman, but we might think his view gives us reason to treat love as an ideal of interpersonal relations. If our task in valuing persons is to be right appreciators, and if in loving a person we appreciate their value most fully, it would seem that loving is the better, the more appropriate, kind of attitude. Now Velleman accounts for the selectivity of love—the fact that we love some people and not others—in two ways. We are, on the one hand, imperfect diviners of personhood, and we are, on the other, finite in our emotional capacities for loving people (p. 372). But why not make it our project to become better at discerning personhood in others, and more capacious in our emotional resources? Since the object of love is personhood, and since personhood is had by all persons, ideally anyone can be an object of love for us. Realising this is a matter of attending well, of really looking at the one before us. We might think Velleman's view invites us to see love in this way as an ideal of interpersonal relations.<sup>98</sup>

---

<sup>98</sup>This is a delicate point. Insofar as Velleman's is a Kantian proposal, love cannot be the maximum response to the value of persons if that means something like: the very best we can do. For Kant, no action is morally good in *excess* of duty for the simple reason that the good just is action from duty. If Velleman's view invites us to see love as an ideal response to the value of persons, and if the view is to remain Kantian, love must be not merely an aim, but requisite for, interpersonal relations; like respect, love must become a *duty* to others. Now Velleman denies that love is a duty to others, but it is not clear why by his own lights it should not be, and it would be a worthwhile further task to see how love may be construed as a duty within a Kantian framework. A natural place to turn would be 'The Doctrine of Virtue' in the *Metaphysics of Morals*. Here we find that "the duty of love for one's neighbor can...be expressed as the duty to make others' ends [one's] own" (MM 6:450). This is the practical love of the 'philanthropist' who shows "active benevolence" towards human beings—"whether or not [she] finds them worthy of love" (6:450). Kant calls this active benevolence, 'benificence': the duty to promote the ends of others.

Velleman is not likely to be sympathetic to the *Metaphysics of Morals* picture, for Velleman does not take love to be in the first place practical—love on his view does not aim for anything (p. 354). Now some of Velleman's antipathy towards conative accounts of love may be mitigated by recalling the reciprocal character of the duties of love and respect in the 'Doctrine of Virtue'. Kant characterises love and respect by analogy with the natural forces of attraction and repulsion: "the principle of mutual love admonishes [rational beings]

### 3.2.5 A Model of Interpersonal Relations

I think there is something deeply attractive about Velleman's conception of valuation, and about the model of interpersonal relations it invites us to hold. In what immediately follows I want to motivate the model by putting before us an intuitive example. The example is an old and oft-cited one: it is of a mother and daughter-in-law from Iris Murdoch's *The Sovereignty of Good* (1970). Let me briefly describe it. There is a mother (M) who is silently disparaging of the woman her son has married (D): she finds her shabby, impertinent, rude—beneath her family's station. But Murdoch stipulates that M is a good woman, she is self-critical and wants to *attend* to D in the right way. "Attention" is a term Murdoch uses for aiming to appreciate in an object of value the value that it really has. So M sets about attending to D, and gradually D is found to be not 'shabby' but 'unpretentious', not 'impertinent' but 'candid', as one set of descriptions is substituted for another.

Murdoch makes the case for this substitution as moral progress, but there is a question about how to understand the progress exactly, and it is important for my purposes to be clear about the moral lesson we are being asked to draw. On one reading of the example, M's work is primarily value-epistemic. She asks herself whether her judgements of D are

---

to come closer to one another"; that of respect "to keep themselves at a distance" (6:449). Though they have opposite valences, love and respect are "united into one duty", in such a way that where the one is the principle of action, the other is "joined to it as accessory" (6:448). So if the maxim of my action is to promote the interests of this man from the duty of love, respect enjoins that I do so without violating his self-respect (6:450). That means that though my action is meritorious—though it puts him under obligation to me—I must make as if I owe it to him. Far from "bristling with uncalled-for impingements" (Velleman, p. 353), Kant tells us we must "carefully avoid any appearance of trying to bind the other" by our beneficence; better still, we should practice beneficence "in complete secrecy" (6:453).

And yet I think no end of secrecy will obscure the fundamental differences between the two accounts. Love on Velleman's view is in the first place a recognition of value, it is not cause for furthering the ends of others. We find, albeit slim, textual grounds for making love in Velleman's sense a duty in the *Lectures on Ethics*. There Kant makes a distinction between well-wishing love—what appears to be the ancestor notion of love as beneficence—and well-liking love—delight in another's perfection (27:417). Both are duties, and the object of well-liking love is the "the kernel of good will" in all persons, including the villain (27:418). Though these are scanty textual grounds they provide a provisional conceptual bridge between Velleman's love of persons and love as a Kantian duty.

justified or not, whether they are after all the product of prejudice: “I am old fashioned and conventional. I may be snobbish. I am certainly jealous. Let me look again” (p. 17). In that case, M’s undertaking is epistemically laudable—she is rooting out false or distorted judgements—and we can understand the achievement as an achievement of right understanding: of forming attitudes and thoughts appropriate to the object of value.

But I take it the example is meant to be one in which M’s initial judgements are not false in this straightforward way. Or at least, we can imagine the case as one in which D *is* rude and shabby in the ways M is initially inclined to find her. How then might we conceive of M’s undertaking? Perhaps we could say that M is engaging in a piece of “constructive emotional reform”.<sup>99</sup> M wishes she did not feel disdainful of D, wishes she could see her in a more loving way, and so because she likes to maintain loving relations with the people in her life, or wants to keep her family together. But now M’s bid looks delusive—she seems to be fudging her new and favourable judgements of D, albeit for laudable ends. While there is nothing amiss in trying to value people, positive valuation tends to resist our efforts at inducement.<sup>100</sup> To understand the example this way would be to saddle Murdoch with an untenable moral psychology, but it would also be to disregard the conditions she lays down for the case: of the various motives we attribute to M we cannot say that “she deludes herself” (Murdoch, p. 18). If she is moved by love, love must have a closer connection to truth, here to the truth about D. “The love that brings the right answer”, Murdoch writes in another connection, “is an exercise of justice and realism and really *looking*” (Murdoch, p. 91). How are we to put these things together?

Velleman himself brings Murdoch’s notion of attention, though he does not discuss the case of M and D, together with Kantian valuation. Both are said to be modes of valuation

---

<sup>99</sup>As David Pugmire puts it (1994, p. 118).

<sup>100</sup>It must happen, as Pugmire (1994) writes “as by grace, or not” (p. 119).

in which what is valued is seen in the right way—*really* seen, where the *really* may be taken to contrast with the merely *apparent*. We might picture Murdoch’s attention, in Velleman’s terms, as our looking through or past a person’s empirical persona and to his true or proper self—to the core of goodness in him. Alighting on that core of goodness our eyes are “opened to what the other really is” (Velleman, p. 362). So M’s love is a discovery, on the one hand, for the object of her love is just what we all possess by virtue of being persons. And M must steel herself to loving, on the other, for on this view we are imperfect diviners of personhood: “[w]hether someone is lovable depends on how well his value as a person is expressed or symbolized for us by his empirical persona. Someone’s persona may not speak very clearly of his value as a person, or may not speak in ways that are clear to us” (Velleman, p. 372). M’s coming to see D as ‘refreshing’ and ‘unpretentious’ is a dawning of personhood, then; ‘refreshing’ and ‘unpretentious’ are what she is valued-as, but importantly, in Velleman’s terms, these are “conduits rather than sources of value” (p. 371). In contrast to the first reading, M’s shift in attitude is not a consequence of eliminating false beliefs or judgements, where D turned out to be not unceremonious (etc.) at all. Rather, as we may imagine, M has looked past D’s irritating habits, past her unceremoniousness, and to Velleman’s true or proper self which is revealed in the attributes marked out by the new thick terms. We could picture this as a shift in perception of what is salient about D. So M’s love is achieved—it does not happen “as by grace, or not” (Pugmire, p. 119); but against the second reading, it is not delusive either. And in this way we can put love together with “justice and realism and really looking” (Murdoch, p. 91).

### 3.2.6 The First Horn of the Dilemma

Let me recall the dialectic. I have been giving an account Velleman’s view of valuing persons. The view says that we are to direct our thought and attention to the intelligible subject, the

bearer of morality and freedom, so that what in human beings is less than ideal becomes irrelevant to how we see them. I have used the example from Murdoch to bring out what I think is attractive in this conception of our attitudes towards persons. I turn now from our attitudes towards persons *simpliciter* to our attitudes towards persons-as-agents-of-actions.

So consider a case in which D wrongs M in some way. She sabotages Sunday lunch, or she gives away the brooch, a family heirloom, M had made her as a gift. And suppose D does this out of insolence or spite. The question I now want to ask is whether Velleman's view invites the consequence that D is not to be held responsible for her actions. The consequence is not entailed by Velleman's view, for his is an account of the attitudes we are to bear towards persons; it is not an account of the evaluative attitudes we bear towards persons *qua* agents of some action. But I submit that it is a natural extension of his view. The extension requires just that valuing persons involves directing our thought and attention only to actions (and to parts of the self) with which the intelligible self is causally involved. Since the sensible self is the locus of actions that fall short of morality, and since the sensible self is not free, if freedom is a condition on responsibility, it follows that a person is not to be held responsible for wrongdoing. In the terms of the example, M is to fix her attention not on D's bad behaviour, but on the person she really is, someone, in Velleman's final gloss on personhood, with the capacity for love of others, even if that is not now exercised. There needn't be revision in M's description of D's behaviour; we needn't imagine that the sabotaged lunch becomes comedy in M's mind, or the dispensing of the heirloom a way to lighten the burden of history; 'sabotage', etc., may be quite the right description. We are imagining, rather, that in alighting on the person, M is alighting on the good in D—on "that better side of the person which constitutes [her] true self" (Velleman, p. 365)—so that D's insolence and spite are consigned to an inessential, or false, part of her nature. It may be that M struggles to maintain this perspective; it may take continual recommitment. But as

Murdoch says, the work of attention is an “endless task”; it requires discipline (1970, p. 28).

Now Velleman may well resist this extension to his view. He may insist on the distinction between the attitudes we take towards persons and the attitudes we take towards the actions of persons, so that the question of whether one should value everyone as a Kantian person is separate from the question of whether one should ascribe responsibility (or not) from the perspective of seeing others as Kantian persons. This is a possible move, though perhaps not in the end a plausible one. For we might think that in taking up evaluative attitudes towards the actions of persons we are always also assuming an attitude to the source of the action, that is, to the person. But it is not my concern to argue this here. What matters for my purposes is not whether Velleman himself is committed to the first horn of the dilemma, but whether some view is which yet provides a defensible model of interpersonal relations. This last I have constructed from resources Velleman brings to the table.

While I have tried to bring out reasons in favour of the view, I want to argue that it licenses revision of our practices of holding responsible for wrongdoing that we have moral reasons to uphold. Consider that the judgement that someone is responsible is a special kind of judgement. This is sometimes put by saying that judgements of responsibility *go beyond* favourable or unfavourable description. In what way they go beyond is a matter of dispute. Some take it to be a matter, in the terminology of P. F. Strawson (1962), of expressing reactive sentiment—generically, praise and blame; others, following T. M. Scanlon (1986), take it to be a matter of demanding justifications, and of engaging in forms of moral argument. It is not important for my purposes which account we adopt. What matters is that judgements of responsibility are constituted by a characteristic range of practical responses to others. These forms of response are a crucial part of our moral lives. That is a broadly Strawsonian point. Strawson took our responsibility practices to be deeply bound up with participation in interpersonal relationships. We are responsive to the good or ill will people bear towards

us and towards other people, just as we are responsive to their attitudes to the principles, values and objects we care about. Our responsiveness finds expression in holding one another responsible when we or others or things we care about are slighted. This is a natural part of our relations with others and it is significant because it communicates our moral expectations to them. In my view the centrality and the role of these practices provides a strong reason to preserve them. Clearly, the moral justification of blame and associated negative attitudes requires extended defence, and in a longer version of this paper I will give that defence. For now, I want to point to just how radical a revision to our practices the Velleman-inspired view requires.

In fact, Strawson's point is taken over and strengthened by Christine Korsgaard (1992), for whom personal relations not only significantly involve the spontaneous exchange of negative reactive sentiment, but actually *require* it—for Kantian moral reasons. So, as we will see, Korsgaard gives us additional considerations for taking the first horn to be unacceptable. Unfortunately, her grounds drive her to the second horn.

### 3.3 Korsgaard on Respecting Others as Agents

Korsgaard begins squarely with the distinction between the intelligible and explanatory standpoints. Her question is whether it is ever permissible to regard a person from the explanatory standpoint, the standpoint from which he is a creature whose behaviour is set by desires and inclinations according to a causal history. Indeed we are often tempted to treat a person who is being infuriating, who is overwrought, enduring difficulties, or is not themselves, as an “obstacle to be worked around” (Korsgaard, p. 315); efforts to engage them rationally give way to a concern with how to manage them best. Korsgaard says she is inclined to think this disrespectful, and for a rationale she turns to Kant:

The reason we must view another as a fellow rational person rather than a psycho-social phenomenon is not that he is *in fact* one of these things rather than another. In fact, he is both. That another is responsible is what Kant calls a postulate of practical reason: a belief or attitude that can be formulated theoretically, but is practical and moral in its basis (p. 321).

And the moral basis of the practical postulate is said to be this:

[T]he moral law [. . .] commands that you treat everyone as an end in himself. Unless you hold others responsible for the ends that they choose and the actions that they do, you cannot regard them as moral and rational agents, and so you will not treat them as ends in themselves (p. 320).

The idea is that treating people as responsible is a necessary condition on our treating them as ends in themselves, and treating them as ends in themselves is commanded by the moral law. One way to make out the connection Korsgaard envisages here between treating someone as a responsible agent and treating them as an end in itself is by noting that, on her view, to treat humanity as an end in itself is to treat someone above all as capable of rational, self-determined choice.<sup>101</sup> Then the thought is that to treat someone as capable of rational choice is to treat them as an agent, and to treat them as an agent is to hold them responsible for what they do (choose). It follows that if we must always treat people as ends in themselves then we must always treat them as capable of rational choice—always treat them as agents. I take it this is Korsgaard’s line of thought. And the conclusion is clear: we must *always* treat agents as responsible, for that is part of what it is to respond in the right way to their humanity. Shifting to the explanatory stance, in which we regard what agents do and say as natural or ‘psycho-social’ phenomena, is tantamount to disrespect.

This position is not without appeal. Part of its appeal is that it requires us to relate to agents as being capable of good will. I am to hold you responsible for the bad choice you

---

<sup>101</sup>For that is what Korsgaard (1996b) takes ‘humanity’ essentially to be. (See especially p. 190). Since Kant does not take the choice between good and evil to be agency in any sense, insofar as Korsgaard’s is a Kantian proposal, the capacity to choose here must be the capacity to choose good ends.



made in acting as you did, because by doing that I show you that I believe you can do better. As Korsgaard (1992) puts it in another connection: holding someone responsible ‘declares to its object greater faith than she has in herself’ (p. 312).<sup>102</sup> It might seem to that extent to join ranks with the Kantian commitment, attractive to many people, that all human beings have moral worth in virtue of their potential to be moral, something it is incumbent on us to recognise by holding them to high standards.

Additional points of appeal may be brought out by noting that, for Korsgaard, to relate to another as a responsible agent is primarily to be prepared to enter into reciprocal relations with him. She says:

To hold someone responsible is to regard her as a *person*—that is to say, as a free and equal person, capable of acting both rationally and morally. It is therefore to regard her as someone with whom you can enter the kind of relation that is possible only among free and equal people: a relation of reciprocity (1992, p. 306).

The idea of reciprocity brings out the egalitarian dimension of Kantianism that many have found attractive.<sup>103</sup> One might think the idea of a reciprocal relation is the idea of a relation in which participants are open to being affected by one another. If we think that includes being affected by one another’s moral point of view, then regarding them from an explanatory perspective will be a way of muting their claim to affect one’s own rational and moral thinking. It is a way of not regarding the person as someone from whom one can

---

<sup>102</sup>Pamela Hieronymi (2001) has argued that to show anger towards a wrong-doer is to give recognition of her moral significance; it is to show that we ‘count her as worth being upset by’ (p. 539). Hieronymi is in line with Korsgaard here.

<sup>103</sup>Though note that insofar as this is meant to be a Kantian view there is a problem with the idea of equality here, for Kant takes the value of human beings to be such that it is non-comparative. This is part of what Kant means when he says that human beings have a dignity and not a price value (4:434-5). And of course being equal is a kind of comparison. Velleman is very clear that the value of human beings for Kant is non-comparative, though he may spoil his insight in saying that “all persons should be judged to have the same value” (p. 367).

learn.<sup>104</sup>

In sum, Korsgaard's claim that we have a moral obligation to hold wrongdoers responsible may be thought attractive because it shows that we see them as capable of good will. It may be thought part of a commitment to egalitarianism, the belief that others are on an equal footing with us, and in that sense able to be our critics—hold *us* responsible—too.

### 3.3.1 The Second Horn of the Dilemma

And yet, for all that, I think Korsgaard is committed to a stark and unattractive view. An obvious problem is that it seems unable to accommodate a certain class of excuses. We tend to take factors like stress, upset, fatigue or grief to mitigate holding people to account. But Korsgaard must say that *not* to do so is to fail to respect their humanity, their capacity for rational choice. This follows from yoking the stance of holding responsible to respect for persons as ends in themselves (given her account of these terms). The result is a rather demanding model for inter-personal relations, one that seems insensitive to a variety of agential frailties.

Now Korsgaard recognises the problem. The rationale she has given on Kant's behalf for holding people responsible seems to entail an "intransigent" view"; this is the second horn of the dilemma:

[I]f we do regard people as free agents, fellow citizens in the Kingdom of Ends, then it seems as if we must treat them as transcendently free and so as completely responsible for each and every action, no matter what sort of pressure they may be under (p. 319).

And this is a *problem* because:

---

<sup>104</sup>Lawrence Stern (1974) nicely makes this point. His term for engaging those we hold responsible in moral argument is 'dialogue', and he sees it as a way of being 'genuinely open to the other's influence' (p. 75).

[T]he obvious fact is that we live in neighbourhoods which are different distances from the Kingdom of Ends, and it seems merciless to give this obvious fact no weight (p. 319).

Indeed so. How does Korsgaard respond to the problem? In essence she responds by appealing to parts of Kant's oeuvre where he seems to take a different view. Specifically, she points to parts of his work where he appears to relax the (would-be) requirement to engage people from the intelligible and not the explanatory standpoint. The claim now is that while it looked as if Kant required that we engage others practically to the exclusion of the explanatory standpoint, that doing so was tantamount to respecting them, in fact Kant mixes these standpoints. So much is revealed by his belief in the efficacy of political institutions, and of moral education, for these require that good conduct can be (naturally) caused (see Korsgaard, 1992, Section V). However much this is true, and however Kant permits a mixing of standpoints, so mixing them is debarred by the rationale for holding people responsible which Korsgaard attributes to him.<sup>105</sup> That it is Kant who asserts *p* and *not-p* is no grounds for concluding that we may after all assert them together.

It is worth considering whether there are better replies in the neighbourhood. To offset at least one form of the intransigence worry, perhaps Korsgaard could argue that the rational-moral capacities constitutive of agency are compromised in circumstances of deep grief, exhaustion or stress. Then when agents do wrong under those circumstances they do

---

<sup>105</sup>It would also seem to be debarred by the position Korsgaard (1996c) takes elsewhere:

[T]he supposed problems about responsibility and ontology arise from a common source: a failure to appreciate the radical nature of Kant's separation of theoretical and practical reason, and of their respective domains of explanation and deliberation. When these domains are separated in the way that Kant's philosophy requires, the problems about responsibility disappear... (p. 160).

The problem about responsibility Korsgaard addresses here is the first horn of a dilemma: that people are only responsible for morally right action. If the separation of the practical and the theoretical is meant to resolve this problem, then it does so only at the cost of landing on the second horn, the problem of intransigence.

not do wrong *as* agents, and they do not fall within the requirement always to hold agents responsible. That requirement would instead read: always hold agents responsible when they act as agents. This might be part of a more nuanced picture of moral capacity that does not see it as possessed in *all* circumstances by those who meet a minimal threshold.<sup>106</sup> But notice that if Korsgaard accepts the amendment, and if, according to what we said above, the capacity to choose rationally which is constitutive of agency is also constitutive of humanity, then it looks as if we have bought a reply to intransigence at the cost of losing grounds for respect. For in falling out of agency, the subject falls out of humanity and thereby insurance by the requirement to treat humanity as an end. This consequence is not one many Kantians are willing to accept, for it undermines a central commitment: that respect is owed to *all* human beings, as Korsgaard (1986b) puts it elsewhere, '[a]s their right' (p. 335). This brings out a deeper problem for the proposed amendment insofar as it is to be a Kantian amendment. And that is that the rational capacities constitutive of agency for Kant are not empirical properties of us, but are borne of our nature as intelligible beings. In that way it does not make sense to think of them as possessed in some but not all circumstances. And that is the reason respect is owed to all rational beings independently of whether their rational capacities are in good form.

A different route for Korsgaard might be to argue that the stringency of Kant's would-be requirement always to hold agents responsible is an *ideal* of human relations proper to inhabitants of the Kingdom of Ends, but not to those in neighbourhoods remote from

---

<sup>106</sup>The minimal threshold view of moral capacity is held in different forms by many compatibilists, and like Korsgaard they seem liable to charges of intransigence. For example, Kadri Vihvelin (2004) defends a view of responsible agency as the possession of a set of intrinsic properties which are the causal basis of a disposition to choose on the basis of reasons. Since the intrinsic properties are possessed by the agent even when the ability is not manifest, say, because she is tired, she is still responsible. Or take the view of Fischer and Ravizza (1998) according to which moral capacity is essentially rational self-control. An agent must be capable of recognising and responding to relevant reasons, a capacity he is said to possess if he is shown to have been reasons-responsive on occasion. As Michael McKenna suggested to me, the intransigence of such views might be mitigated by taking the causal basis of agential capacity to be affected by external factors.

it.<sup>107</sup> Given the conditions of the actual world, where agents are inconstant, imperfectly rational and suffer defects of natural constitution, we need the governance of a non-ideal theory. Then the ideal theory becomes not an obligatory standard for inter-personal relationships, but a goal they should strive to meet, or a model of what they would ideally be (p. 342-3). The obvious reply to this line of thought is that in the Kingdom of Ends there would *be* no wrongdoing, for everyone would be perfectly rational, so there would be no holding responsible for wrongdoing.<sup>108</sup> As an ideal model of inter-personal relations Korsgaard's rationale is incoherent, and since it fares badly as a model for current relations, by Korsgaard's own admission, it is unclear what end it serves.

## The Diagnosis

I began this paper by considering Kant's dual conception of human beings in the *Groundwork* as it is given in the language of standpoints. Kant thinks, as we saw, that we take a standpoint on ourselves as intelligible beings from the first person perspective of deliberation. And Kant thinks, as we saw, that this standpoint excludes a standpoint on ourselves as sensible beings, for as a matter of practical necessity, the thought was, we cannot both deliberate and take our deliberation to be under causal influence from nature—from our nature as embodied creatures. The two standpoints were in this way shown to be exclusive

---

<sup>107</sup>I adapt the following line of thought from the style of reply Korsgaard (1986b) gives to the problem of Kant's uncompromising idealism in the face of the contingencies of a non-ideal world. These include Kant's seeming injunction against lying to the murderer at the door, his injunction against suicide in the face of acute misery, and his requirement to perfect our talents even when we lack the means or opportunity to do so. To respond to these problems Korsgaard adapts John Rawls' distinction between ideal and non-ideal theory, what Korsgaard calls a double-level theory, to Kant.

<sup>108</sup>Korsgaard shifts between conceiving of holding responsible as a stance towards someone's actions or attitudes, and conceiving of it as a stance towards persons themselves, wherein we speak of so-and-so's being a responsible person. It makes sense to say that in the Kingdom of Ends, relations will be of a reciprocal kind among responsible people, but this is not the sense of responsibility that needs to be addressed here, which is responsibility for some action.

for Kant.

But then the question was whether we are to take a perspective on *other people* as intelligible beings, and whether in taking that perspective we are to exclude considerations about their sensible nature from practical ways of relating to them. I offered the positions of Velleman and Korsgaard as contemporary Kantian positions which in different ways take over the transposition of the intelligible standpoint to the third-person case. Both take regard of others as intelligible beings to be borne of Kant's theory of valuation, in particular, of what it is to value persons as ends-in-themselves. For Korsgaard, as we saw, to value persons as ends-in-themselves is to assume the intelligible standpoint on them, and that means to treat them always as agents. To take up the explanatory standpoint, to regard them as sensible beings, would be to fail to keep "the humanity of others and so their capacity for good will always before our eyes" (Korsgaard, p. 308); and more strongly, it would be to "write [someone] off as a person" which is something we do only "at [our] own moral peril" (p. 314). On the other hand, for Velleman, as we saw, to value persons as ends-in-themselves is to appreciate the real or proper part of them which is the purely intelligible part. Sensible considerations are excluded from valuing the real and proper part inasmuch as they belong to the person as he merely appears—to what Kant calls the human being, and Velleman the "empirical persona" (Velleman, p. 371).

My strategy in the paper was to argue against these versions of Kant's theory of valuation by showing how they entail, or in Velleman's case invite, unpalatable consequences for our relations with other people. In particular, I argued that they deliver peculiar results for holding others responsible for wrongdoing. They deliver the result either that people are never to be held responsible for wrongdoing or that people are always to be held responsible for wrongdoing. My argument against these conclusions turns on offering a normative justification of our practices of holding people responsible. Against the Velleman-inspired

view I argued that holding others responsible for wrongdoing is sometimes appropriate. I argued that the practical forms of response constitutive of holding responsible—making demands for justification, engaging in robust forms of moral argument, blaming and related forms of censure—play an important role in communicating our moral expectations to others, and in holding them to standards we think they should share. Against Korsgaard, on the other hand, I argued that holding others responsible for wrongdoing is not always appropriate. I argued that our practices of excusing wrongdoing agents whose behaviour is borne of stress or ill-health, or whose behaviour is out of character, are appropriately excused on occasion. In this I take myself to be defending our ordinary practices of holding people responsible for which it is sometimes but not always appropriate.

One way to put my criticism is to say that the Velleman and Korsgaard style views invite overly simplified accounts of the conditions of responsibility. To the question of under what conditions we are to hold another responsible for wrongdoing they give brute answers: it is never for the one, and always for the other. And that is because they take all that is relevant to determining those conditions to be whether someone possesses a rational nature. If they do then Korsgaard says that they must always be held responsible for their conduct—to hold them responsible is what it is to value their rational nature. If they do then the Velleman style view says that they must never be held responsible for their conduct—to not hold them responsible is what it is to value their rational nature. My criticism may be put by saying that to work out the conditions under which someone is responsible, and so how we are to relate to them, we need to know more than whether they possess rational nature—we also need to know empirical facts about them. If a person has been impertinent and ungracious, then we need to know this. If he has been suffering under strain or exhaustion then we need to know this. The problem is that the views take considerations about another's sensible nature to be irrelevant. But they are not: our stance towards others is and should be sensitive

to empirical matters.

### 3.4 A Disjunctive Conclusion

What generates the unpalatable consequences for responsibility, then, is a conception of human beings according to which there is the part which engages our practical responses, rational nature, and the part which engages our theorising, sensible nature. But sensible nature does not come into view as practically relevant, and agency does not come into view as explanatorily relevant. Against this, and on the basis of the line of criticism developed here, it seems to me that there is one of two positions available. Let me outline them in turn.

#### 3.4.1 Concessive Kantianism

The first is a concessive Kantian position. It retains a dual conception of our nature as intelligible and sensible beings, and it says that a person's intelligible nature is the object of valuation. But it adds that to appropriately value another's intelligible nature is not to exclude consideration of sensible nature. On this sort of position, consideration of sensible nature is part of what it is to value intelligible nature. So while on the view I attributed to Velleman appropriate valuation of the intelligible subject meant discounting the infractions borne of sensible nature, on this view counting them would be precisely to value the person appropriately. Engaging the person in robust forms of moral argument, demanding justification and so on would be what it is to respect them. In this the view would be closer to Korsgaard's position. But it would be unlike Korsgaard's position in that not taking account of the frailties borne of an empirical constitution would amount to improper valuation of the intelligible subject. The form of intransigence that Korsgaard is forced into by her injunction always to regard persons as agents would on this view be a



failure of respect.

The question is whether this is viable as a Kantian position, or whether it is ruled out given attendant Kantian commitments. One might think it ruled out, for example, on grounds of Kant's rationalism. Kant takes moral requirements to hold for all rational beings as a matter of necessity, and since he thinks that necessity cannot be grounded empirically, he excludes empirical considerations from the grounds that can be given for moral principles. It is in this sense that he takes himself to be doing metaphysics of morals. Now one formulation of the supreme principle of morality is that we must value persons always as ends-in-themselves, and we have seen that that may be understood to mean that we must value persons as intelligible beings. But does it imply further that in valuing persons as intelligible beings we must exclude consideration of their sensible nature? Does Kant's rationalism license this exclusion? Now in a clear sense, for a morality that is meant to hold of all and purely rational beings, considerations about sensible nature will not be forthcoming. Purely rational beings do not do wrong, and do not have bodies, so we need not be told how to treat them when they do wrong, or when their bodies impede their agency. But human beings are not purely rational beings, we are taken to have a dual nature. How does Kant's rationalism speak to us? One way to focus the question is to ask whether Kant includes in his formulation of the principle of morality considerations that are particular to beings with a dual rational-sensible nature. He seems to do precisely that in the *Groundwork*. The bulk of the *Groundwork* addresses itself to beings with a dual nature. This is why the supreme principle of morality is formulated as an imperative—a perfectly rational will does not need to be commanded to do his (its) duty. This is also why the deduction of morality in the third part of the *Groundwork* aims to show that the categorical imperative is valid for and binding on human beings. If this is right then it is not clear that Kant's methodology in the *Groundwork* would give us grounds against the concessive

Kantian position.

### 3.4.2 Rejection of the Distinction Between Intelligibility and Sensibility

A second and non-Kantian position says that we ought to give up on a conception of human beings as having a dual intelligible-sensible nature. On this view we are mixed beings, we are sensible agents, and should be treated and understood both ways together so that our sensible nature is practically relevant, and so that our agency is a sensible feature of us.

I want to offer reasons in favour of this kind of position by turning to a neglected observation of P. F. Strawson (1962). Strawson gave analogues, albeit metaphysically deflationary analogues, of Kant's distinction between the two standpoints; Strawson's were the participant stance and the objective stance.<sup>109</sup> But what interests me is a move Strawson makes towards effacing the distinction he draws for us. "I am aware", Strawson writes, that in distinguishing the two stances "I have presented nothing more than a schema, using sometimes a crude opposition of phrase where we have a great intricacy of phenomena" (p. ). He goes on to describe some of that intricacy by discussing cases in which we do not take either the participant or the objective stance in any pure form. Strawson's example is of raising children. Children are creatures in the process of acquiring the capacities that make them appropriately held responsible, and of the capacities that allow them to appropriately hold others responsible. But children are not yet fully possessed of either capacity. So our treatment of children represents what Strawson thinks of as a "compromise" between the two stances; we shift back and forth between them, sometimes assuming the one stance and

---

<sup>109</sup>The participant is the stance we take towards agents—those we engage in the practical forms of response constitutive of holding responsible. The objective is the stance we take towards non-agents—those we withdraw from the measure of our normal practical responses in favour of a stance of prediction, explanation and control. So as Strawson sets up the distinction, we take the two stances towards different kinds of being, agents and non-agents. In this Strawson's differs from the Kantian distinction which imagines that we are both kinds of being—that we take up both positions, as it were, for different purposes.

sometimes the other.

I want to suggest that most of us represent intermediate cases in Strawson's sense, and I want to suggest that our intermediate position should make us think differently about the distinction he draws. To motivate my claim, let me adduce some very general considerations about human beings, and I apologise in advance both for stating what may seem like crude platitudes, and for stating them in a way that lacks much necessary refinement and complexity. Adult human beings are in various stages of passage to rational agency, and the passage is not linear—we often regress and then regain ground. In this passage our rationality is continually hostage to our physical life, to low-level physiological matters like how much sleep we had and when we ate, to more psychological matters like how concentratedly we are attending to our environment, or how anxious or confident we feel. Our rationality is hostage to our environments too as, in a very deep way, our environments shape our upbringing with all the accidents of education and parentage that brings, to more fleeting arrangements like the noise outside our window or even our position in a room.

Now for Strawson our attitude towards normal adult human beings is in the normal case an attitude towards rational agents. Strawson considers that we sometimes take up an attitude towards normal adult human beings in which we regard them as patients—as merely sensible creatures. Then we shift to the explanatory standpoint. We may do this, Strawson thinks, for a variety of reasons: 'as a refuge, say, from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity' (p. ). And he thinks that we may vacillate between these perspectives, taking up sometimes the one and sometimes the other. In this way our attitude towards normal adult human beings is sometimes like our attitude towards children in being intermediate between the two stances. But note how different that idea of vacillation is from what I take to be a more standard kind of case, that of excusing someone for, say, forgetting about a date with us because they were preoccupied

with their work. It does not seem to me that in excusing a person under these circumstances we switch from regarding them as an agent full-blown, to regarding them as a creature full-blown, albeit in such a way that we flicker back and forth in our regard. Rather, it seems to me that our attitude is towards an embodied agent, someone whose rationality is hosted by and hostage to the physical world. He is for us neither a pure agent nor a purely causally determined one.

I am willing to grant that the distinction Strawson draws may be useful for certain purposes, but I want to suggest that we should understand it to be quite porous. That is, if we take up a perspective on someone as an agent, then we should do so in a way that is informed by a conception of them as a sensible being. Likewise, if we take up a perspective on someone as a sensible being, then we should do so in a way that is informed by a conception of them as an agent. My suggestion is that the stances ought not to be thought of as isolated from one another. It is perhaps worth recalling here that it is just this mixed perspective that Korsgaard wants in the end in order to accommodate a variety of agential frailties. Now in a clear sense Korsgaard is prevented from taking this mixed perspective for reasons Strawson is not. For Korsgaard takes her analogue of Strawson's participant stance to be morally required towards beings with a rational nature. So where Strawson says that we take the participant stance but shift also to the objective stance in such a way that our stance may not be purely one or the other, Korsgaard must say that to so shift is to fail to treat a person as an end-in-itself. Her position in this way forces her to insist on a thorough screening off of the two perspectives from one another. And that means that the friend who forgets our arrangement cannot be excused on grounds of strain or stress. Excusing him would involve an illicit mixture of the two perspectives. But I want to add to this now that Strawson's idea of two perspectives as separate but potentially held together in a kind of tension, is not sufficient either. If it is to be adequate to our perspective

on normal adult human beings, the two sides of the distinction must shape one another and not merely be flickered between.

Now we said that Strawson's distinction is a metaphysically deflationary one, but the distinction (or version of the distinction) in Kant is metaphysically robust. In Kant we are thinking not of ways to regard someone, but of the positions they occupy as members of different worlds.<sup>110</sup> Our rational agency on this conception is really not an empirical feature of us. In this way, the distinction seems less amenable to softening. And if the above considerations are on the right lines, if the rational nature of human beings should be thought of and engaged as part of their sensible nature, then it might seem that we lose our grip on the Kantian picture of a dual-natured rational and empirical being.

---

<sup>110</sup>Note that Korsgaard reads Kant's two standpoints in a non-metaphysical way in line with the two-aspects reading of Kantian idealism. In this she is able to see herself as an inheritor of a version of Strawson's distinction. But see footnote 1 above.

## 4 On Regress Arguments for the Value of Valuers

### 4.1 Introduction

The topic of this paper is, in a sense to be qualified, the value of human beings. More particularly, I am interested in examining accounts on which human beings are *valuable simpliciter*—valuable independently of whether we are *valuable for* anything or anyone. More particularly still, I am interested in examining an influential argument schema for the value simpliciter of human beings. The argument schema, and it is a *regress* argument, has two parts, and is simple to state:

- (i) for anything to be valuable, there must be something which simply is valuable—which is valuable simpliciter; the chain of dependence between values must come to an end; (ii) human beings meet the criteria for being valuable simpliciter.

Versions of the argument have been put forward in discussions of human value by Christine Korsgaard and Joseph Raz. But part (i) of the argument has rather wide currency. Versions are given by G. E. Moore, and Aristotle; and an analogous argument is made in epistemology concerning, not value, but justification and knowledge. The conclusion of my paper is in a clear sense negative: I will argue that both versions of the schema fail to establish their conclusions. But there is a positive lesson. The positive lesson is that we have reason to develop an account of the value of human beings without making recourse to the notion of value simpliciter. We can capture core desiderata for an account of the value of human beings by taking all value, including human value, to be value for something or someone.

It is a background hypothesis of this paper that there is no such thing as value simpliciter. My background hypothesis is that all value is value for something or someone, or more strongly, that value *is* value for. By that I mean that when something is valuable it is valuable

*because* it has some feature or features which make it such that it is valuable for something or someone. Now I do not propose to defend this background hypothesis directly here. Rather, my strategy is to argue against the view that we *must* take something, human beings, or valuers, to be valuable simpliciter for anything to be valuable at all. The hope is that in securing this more restricted conclusion, an important step is made towards defending the background hypothesis.

The plan for the paper is as follows. In Section 4.2 I explicate the notion of value simpliciter, distinguishing it from related notions. In Sections 4.3 and 4.4 I give the two versions of the argument schema for the value simpliciter of human beings. In Section 4.5 I lay out various alternatives to the claim that the chain of dependence between values must come to an end with value simpliciter. In Section 4.6 I give reason to favour the alternative on which the regress terminates in a kind of reflexive relation, so that objects and activities of value are valuable for human beings, and human beings are valuable for ourselves.

But first a note about terminology. I have been speaking of the value of ‘human beings’, but there is a tradition, and it is part of a rationalist tradition, of speaking rather of the value of ‘persons’. For Kant, human beings have an empirical and a noumenal nature and it is our noumenal nature—what, as *persons*, we share with rational aliens and angels—which makes us valuable. In speaking of the value of ‘human beings’ I am marking a departure from the Kantian tradition. For I am interested in the value of human beings as we are actually, that is empirically, constituted.<sup>111</sup> But talk of human value is not without problems of its own. For one, as Harry Frankfurt (1971) has made the point, it carries the implication that we are interested in the value of human beings *as opposed to* other animals. And yet, we would be interested in whatever feature makes us valuable whomever possessed it. It is perfectly *possible* for other creatures to possess the feature in question, even if we seem to be

---

<sup>111</sup>I give reasons for this background commitment in ‘Responsibility and the Value of Intelligible Beings’.

the only creature to possess it in fact.

Another problem stems from well known difficulties with giving a unified account of the value of human beings. The value of someone in a persistent vegetative state is not well accounted for in terms of agency, or reasoning, or valuing—or other common candidate bases; and likely not a capacity for these things either. Their value, and the obligations we have to treat them in certain ways and not other ways, will need to be grounded differently.<sup>112</sup> For this reason, we might speak not of the value of ‘human beings’, and not of the value of ‘persons’, but of the value of ‘agents’, or ‘reasoners’, or ‘valuers’. Since I am principally concerned with views which treat our capacity to value as the basis of our value, I will speak of ‘valuers’. If I use the term ‘human being’, it should be kept in mind that I am concerned with the value of human beings as valuers.

## 4.2 The Notion of Value Simpliciter

I turn, then, to the notion of value simpliciter which is at stake in the arguments under consideration, and begin by offering a working definition. Something is valuable simpliciter *iff* it is valuable independently of whether it is valuable for anything or anyone. Rough synonyms for value simpliciter include ‘absolute value’, ‘value sans phrase’, and ‘value in itself’. Though there are conceptual differences, these are all terms which mark value, or goodness—terms I will use interchangeably—as non-relational.

The opening of Kant’s *Groundwork* (4:394), where Kant is discussing the value of a good will, helps to unpack the notion. It is not important for our purposes that Kant is discussing

---

<sup>112</sup>In stating my preference for a pluralistic account of the value of human beings, I am committed: (i) to giving some basis or criterion for the value of human beings; (ii) to taking that basis or criterion to be something other than that we are a certain kind of being; (iii) to making the practical significance, and so the value, of a given criterion clear. These commitments would take some argument, but I leave that task to one side here.



the value of a *good will*; what matters is his characterisation of its value features.<sup>113</sup> Kant tells us that a good will is not good because of what it effects or accomplishes, or because of its suitability for attaining some end. Usefulness, Kant says, would add nothing—it would serve only to attract those who have no appreciation of its proper worth; and fruitlessness would take nothing away. As the oft-cited passage would have it:

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 the greatest efforts it should yet achieve nothing and only the good will were left [. . .]—then, like a jewel, it would still shine forth by itself as something that has its full worth in itself (4:494).

To capture what is distinctive about the notion, it is worth considering a contrast class. So compare our usual ways of explaining the value of things—the value of a pencil, or a gift, or a trip to a museum. We tend to explain the value of these things by pointing to a positive difference they make to something or someone: they are useful, or beneficial in some broader way. The pencil allows us to make marginalia, the gift to show gratitude for hospitality, the trip to a museum to bring our studies of Tutankhamun to life.<sup>114</sup> The example of the pencil is an example of instrumental value, of something that is valuable because of what it brings about, or causes. The examples of the gift and the trip to the

---

<sup>113</sup>Note that Kant characterises the good will by way of more than one value notion. The good will is absolutely valuable—valuable whether or not it is valuable for anything or anyone. This is what I am calling value simpliciter, and it is the notion I am focussing on. But the good will is also unconditionally valuable—roughly, valuable in all circumstances. These concepts are related in Kant, but I am not here drawing attention to the would-be unconditionality of the value of the good will.

<sup>114</sup>At least, these are explanations of the value of certain objects and activities of value along one explanatory axis. And there is likely more than one. For example, we sometimes seek to explain the value of things by pointing to the features on the basis of which they are good for something or someone. For example, having friends is valuable because we are known by our friends, can express ourselves uninhibitedly with them, can count on them for advice (and so on), where these things are valuable because they are valuable for us—perhaps we think they are part of what make our life go well. I distinguish these axes of explanation in ‘On the Value of Human Beings’, but for our limited purposes here we can leave the complexities aside. I note however that drawing a distinction of this kind would be part of a response to the challenge to the good for theorist raised by Wolf (2011).

museum are examples of non-instrumental value. The gift does not cause but is expressive of gratitude, and the trip to the museum does not cause but is part of the enlivening of one's studies, where gratitude and enlivened studies are valuable for us, as it were, for their own sake. So we have two kinds of value for something: value for the sake of something else (instrumental value), and value for its own sake (non-instrumental value). Crucially, the kind of value at issue in the opening of the *Groundwork* is wholly unlike either case. It is the kind of value that is valuable whether or not it is useful or more broadly beneficial for anything *at all*.

Kant seems to see, rightly, that there is something striking about the idea.

There is, however, something so strange in this idea of [...] absolute worth [...], in the estimation of which no allowance is made for any usefulness, that, despite all the agreement even of common understanding with this idea, a suspicion must yet arise that its covert basis is perhaps mere high-flown fantasy... (4:394).

If it is liable to be thought high-flown fantasy, we might ask, why covet it? I take it there are two related intuitions here. First, the notion seems to bear out the distinctiveness of the value of human beings. We have the intuition that human beings are valuable whether or not we are valuable for anything or anyone else. Our value is thought to be in this way *non-derivative*. To take human beings to be valuable simpliciter is to capture the intuition about non-derivativeness by invoking a distinctive kind of value, one that yields an explanation that is quite unlike the explanation of other valuable things. Second, the notion seems to bear out the distinctiveness of the *duties* owed to human beings, the sense in which our duties regarding them are not defeasible. The notion of value simpliciter would seem to capture the distinctiveness of these duties by invoking a distinctive kind of value, one that has a practical valence that is quite unlike the practical valence of other valuable things.

Now, for Kant, the value of human beings, or rather, the value of *persons*, is different in

kind from the value of other kinds of things. Persons have a value like nothing else in the world. The views I will engage with in this paper share that intuition: the value of valuers is different in kind from the value of other valuable things, in being valuable simpliciter. But it is worth pointing out that there is a long tradition of invoking the notion of value simpliciter for other valuable things. For figures in the tradition of G. E. Moore and W. D. Ross, candidate bearers of value simpliciter include aesthetic appreciation, knowledge, and virtue.<sup>115</sup> On these views, the appreciation of aesthetic objects is valuable independently of whether it in some way benefits those who engage in it, by, say, enriching their lives. The thought is rather that value dwells in the aesthetic participant.<sup>116</sup> Likewise, having knowledge is thought to be valuable independently of whether it expands the imagination, permits good deliberation, enlarges one's sense of the possible, or in some other way enriches a person. So nothing about the concept of value simpliciter requires us to restrict it to human beings, or more generally, to valuers.

It is worth pausing at this point to situate the notion of value simpliciter in relation to other philosophical topics in the theory of value. Recent discussions of the contrast between value simpliciter and value for, or as it is also put, between good and good for, address a particular metaphysical question: is good a relational property, or is it non-relational.<sup>117</sup>

---

<sup>115</sup>For Ross (2002, pp. 65-7) the list includes pleasure, knowledge, and virtue; for Moore (1903) the greatest goods fall under aesthetic enjoyment, and the pleasures of human intercourse (p. 203 and passim).

<sup>116</sup>For a recent defense of the Moorean position, see Regan (2004).

<sup>117</sup>At least, this is how I would phrase the issue. For recent work on the good as good for see Kraut (2007, 2010, 2011); Rosati (2006, 2008, 2009); Smith (2003); Regan (2004); and Raz (2001). Kraut gives the fullest treatment. I am sympathetic to many features of his account, though I do not share his view that the issue turns principally on whether value is taken to be a reason-giving property or not, so that "friends" of value simpliciter see value as reason giving, and "foes" do not. See Kraut (2011, pp. 16-7, and passim). That makes the question indistinguishable from the question of whether a buckpassing account of value is true. But in my view the issue about buckpassing is orthogonal; it comes up in giving a substantive meta-ethical theory of value. Moreover, it seems to me that one can think there is such a thing as value simpliciter, without taking value simpliciter to be reason-giving. It can be taken to play a different theoretical role, for example, an explanatory role. Finally, I do not share Kraut's eliminativism about value talk, and *a fortiori*, about talk of the

This question is sometimes put in semantic terms: is ‘good’ a relational predicate—such that it functions logically as good for some S—or is it non-relational.<sup>118</sup> The question has a long history, but it departs from the focus in meta-ethics on the question of whether evaluative judgements are true or false, and whether value predicates refer to real features of the world. It departs from these questions by focusing first—before one turns to questions about cognitivism/non-cognitivism, and realism/anti-realism—on a specific metaphysical issue, namely whether value-properties are relational or not. The notions of value simpliciter and value for are compatible with a range of substantive meta-ethical views, and the question of interest here does not require one to take a stand. Meta-ethical questions come up in an account of the nature of value only in a second step, and this is not a step taken in the present paper.<sup>119</sup>

Partly for this reason, it is important to head off any unclarity introduced by the jewel analogy in the quotation from Kant above. There are difficult questions about how to interpret the analogy, but it can seem to encourage the view that absolute value, or value simpliciter, is a self-standing value property, of the sort posited by value realists. Then there are questions about how Kant can accommodate that kind of picture given his broader philosophical commitments. I do not wish to enter upon those controversies here.<sup>120</sup> The relevant question for my purposes is not whether Kant can be a moral realist or not, but

---

value of human beings. See for example Kraut (2011, p. 40, f.n. 1). I think it is fine to call objects, individuals, or activities ‘good’ or ‘valuable’, and fine to say that because they are good or valuable they give us reason to respond to them in certain ways. All this is consistent with taking the objects to be valuable because they are valuable for something or someone. On this last issue, see my ‘On the Value of Human Beings’.

<sup>118</sup>For relevant work on the semantic side, see Geach (1956), Foot (1985), Thomson (2001, 2008). I agree with Kraut (2011, p. 28) that the question is not best construed as being about the *meaningfulness* of sentences with the word ‘good’ in them, but of understanding what makes those sentences true.

<sup>119</sup>I am grateful for discussion with Katja Vogt and Ralf Bader on this point.

<sup>120</sup>For a non-realist reading of the jewel analogy, see Sensen (2009a,b).

whether he is committed to a notion of value simpliciter, something's being valuable whether or not it is valuable for anything or anyone. I take it the answer to that question is clearly—yes. It is then a further matter whether what Kant calls absolute value is a self-standing value property, or something that categorically must be valued, so that “X has absolute value” is equivalent to “X is such that it must be valued no matter whether anyone wants to value it; no matter whether valuing it is useful to anyone, or beneficial; indeed, no matter whether it makes a difference of any kind to anything or anyone at all”. If the second reading is to be preferred, then the philosophical burden, for Kant, shifts from postulating a self-standing value property to postulating categorical bindingness.<sup>121</sup>

There is a question of whether the notion of value simpliciter is the same as the notion of ‘intrinsic value’. But that notion is ambiguous. It can be taken to mean, simply, valuable for its own sake; or it can be taken to mean, more ambitiously, valuable in its own right, irrespective of whether it is valuable for anything else. Taken in this second way, the notion is comparable to the notion of value simpliciter which interests me. But the notion of intrinsic value typically goes together with a host of other ideas. For example, a marker of intrinsic value has traditionally been the so-called ‘isolation test’ described by Moore (1903), wherein we ask whether something of value would be valuable if it were the sole thing in the universe. The isolation test is too restrictive for the notion of value simpliciter. The notion of value simpliciter may be such that other things must be good for it—meaning that it cannot exist in a world taken in isolation—though its value does not depend on its being good for anything or anyone.<sup>122</sup>

---

<sup>121</sup>I use the term ‘postulation’ advisedly, here. On either reading, it is best to take Kant to treat the matter not as a subject of proof, but of postulation. Incidentally, this is a way of overcoming some of the epistemological problems Sensen (2009a,b) raises for the realist interpreter.

<sup>122</sup>This is the notion of interest for Raz (2001, pp. 149-151), who makes the case that value simpliciter need not be ‘autarchic’.

Or again, it is generally thought that what is intrinsically valuable is valuable on the basis of its intrinsic properties. But I take it that it is perfectly conceivable for something to have value simpliciter on the basis of its extrinsic properties. For we should draw a distinction between the nature of the properties on which something of value depends—whether they are intrinsic or extrinsic properties; and the explanation of what makes those properties valuable—whether their value is to be explained in terms of being good for something or someone, or whether their value is to be explained in terms of their being valuable in their own right, or simpliciter. To bring out the distinction, consider that we might take a work of art to be valuable on the basis of its intrinsic properties—its use of colour and light etc.—though we take its value is to be explained in terms of its capacity to enrich our lives in certain ways—in terms of its being non-instrumentally good for us. Then something can be valuable on the basis of its intrinsic properties, though its value is relational. Or again, we might have a view (however implausible) which says that the oldest tree in the world is valuable in virtue of being unique or rare—where uniqueness or rarity are extrinsic properties—though the value of something with those properties is thought to be self-standing: it is valuable whether or not it is valuable for anything or anyone. Then something can be valuable on the basis of its extrinsic properties, though its value is non-relational. The view need not be plausible; conceivability is all we need to see that the questions come apart.<sup>123</sup>

For these reasons, while the notion of intrinsic value lines up with the notion of value simpliciter in some of its uses, since ‘intrinsic value’ is ambiguous, and since it is encumbered with historical-cum-conceptual baggage that need not encumber the notion of value simpliciter, I will not make use of it.

---

<sup>123</sup>Variants of this distinction, and of cases which purport to bring it out, are discussed by Kagan (1998) and Rabinowicz and Rønnow-Rasmussen (2000).

### 4.3 The Regress Schema: A Source Intuition

I turn, then, to the first of two versions of the argument schema for the value simpliciter of human beings. Both versions rest on an intuition about the structure of value explanations. The first starts from the thought that value has a *source*—that it must *come from* somewhere. A version of the argument is given in Korsgaard (1983, 1986a). For ease of exposition, I will take it to proceed in two stages.<sup>124</sup>

#### 4.3.1 Stage I

Stage I bears comparison with the regress argument for the existence of a first cause.<sup>125</sup> It begins with the thought that some things are good, or have value, and it asks: what is the *source* of the value of these things? Where does their value *come from*? The conclusion of the argument is that their value must ultimately come from something that has value in itself, or simpliciter. The argument is that without a source that is valuable simpliciter, we would have to take the value of objects or activities to derive from other things, which

---

<sup>124</sup>There are actually two versions of the argument schema in Korsgaard's presentation which are not clearly distinguished, an argument from justification, and an argument from explanation. I focus on the latter, and I abstract away from various subtleties in Korsgaard's presentation. Korsgaard takes her argument to be an interpretation of Kant's argument for the value of humanity, but it is not best so construed. [See Timmermann (2006)]. Note that where I am using 'value simpliciter' and 'value for', Korsgaard makes use of a contrast between 'unconditional' and 'conditional' value, one she takes to be equivalent to the contrast between 'absolute' and 'relative' value in Kant. I would argue that these are actually different distinctions (the conditionality/unconditionality distinction pertains rather to the circumstances under which something is valuable). The use of different terminology has one main consequence for my purposes. The way in which value can be relational, on Korsgaard's account, is that it *derives* from valuers. This is different from value being relational in the sense that it is valuable-for something or someone—useful or more broadly beneficial to it or them. Though it is a key difference from the notion of issue in Kant, and different from the notion at issue in Raz, it shares sufficiently much with them to warrant consideration here.

<sup>125</sup>Korsgaard invokes Kant's argument for a first cause from the antinomies of the first *Critique*: "if we explain a thing in terms of its cause, we then go on to explain the cause itself in terms of its cause, and this process continues. Reason does not want to rest until it reaches something that needs no explanation [. . .] say, something that is a first cause or its own cause" (1983, p. 259). Though the first cause argument leads to an antinomy in theoretical reason, Korsgaard argues that there is no such antinomy in the argument for the existence of value simpliciter (what she is calling 'unconditional value') in the practical case (1986, p. 119).

derive their value from other things in turn, and so on ad infinitum.<sup>126</sup> But then it is hard to see how value comes on the scene, or gets going. So we must posit an original source of value—something that is valuable simpliciter. As Ross (2002, p. 75) encapsulates the thought: if nothing had value in itself, “value would seem always to be borrowed, and never owned; value would shine by a reflected glory having no original source”.<sup>127</sup>

#### 4.3.2 Stage 2

Stage 2 of the argument says that human beings, as valuers, are the source of the value of other things. Human beings are the source of value because we confer value on things by choosing them according to rational standards. The capacity to choose ends according to rational standards is what our humanity consists in. And our humanity is said to be absolutely valuable—valuable simpliciter.<sup>128</sup>

#### 4.3.3 Responses to the argument

What are we to make of the argument? Of the first stage, we might urge that the idea that value has a *source* is a supposition of meta-ethical constructivism, and is denied by realists.<sup>129</sup> To that extent, the argument rests on a contentious, theoretical premise. But perhaps more importantly, even if one thinks that value has a source, it does not follow that the source is itself valuable, let alone valuable simpliciter. For one might think that value ‘comes from’ valuers who are themselves not valuable, or who are valuable only insofar as they are a necessary condition on the value of other things. To think otherwise is to be misled by a

---

<sup>126</sup>The chain of derivation Korsgaard considers, and she is here invoking candidates which Kant discusses, are from our ends, to our desires, to the faculty of desire, to reason.

<sup>127</sup>Korsgaard (1983) also quotes Ross, at p. 262.

<sup>128</sup>See Korsgaard (1986a, pp. 123-4).

<sup>129</sup>By for example Raz (2003, p. 140, fn. 15). See also Scanlon (2003).



picture of transference, a picture encouraged by talk of value ‘traveling’ from human beings to objects (Korsgaard 1983, p. 261), as if value were an electrical charge, or a substance that rubs off on things.<sup>130</sup>

Of the second stage of the argument, we might agree that the capacity to set ends, or as Korsgaard also puts it, the capacity to value, is valuable. But why think it is *absolutely* valuable—valuable simpliciter? For Kant, our humanity is absolutely valuable not because it is a capacity to set ends in general, but, if we are to make use of this idiom, because it is a capacity to set *moral* ends. It is our ability to align ourselves with our rational as against our sensuous nature that gives us special worth. If we give that up, as Korsgaard gives it up, then it is not clear why the power to set ends should be taken to call for a special kind of value.<sup>131</sup> An alternative story would take the power to set ends, or as I would prefer, the power to have or pursue ends, to make us valuable because it makes us good for something. As we will see, I think something close to this is likely to be true.

#### 4.4 The Regress Schema: An End to the Chain of Dependence

Now consider the second version of the argument. A version is given in Raz (1986, 1997, 2001). Like the first, it comes in two stages, and like the first, it rests on an intuition about the structure of value explanations.

---

<sup>130</sup>This general line of argument was made by Schneewind (1998), Sussman (2003, p. 352) and others, and was conceded by Korsgaard herself. In an appendix to (Korsgaard 1996a), she writes that whereas in her early writings she took humanity to have value simpliciter, in the later writings human beings confer value even on themselves.

<sup>131</sup>On this point, see Timmermann (2006).

#### 4.4.1 Stage 1

Stage 1 starts with the thought that the value of many valuable things is explained in terms of their being valuable for something or someone. In Raz's example, using engine oil is good because it allows the engine to run more smoothly. Cinema going is good, as friendship and games of tennis are good, because they are good for those who participate in them (to the extent the movie is worth watching, the friends are worth having, and each is engaged with in the appropriate way). Next, it is noted that when we explain the value of these things, we suppose that the thing for which they are good is itself good for something or someone in turn. Engine oil is good because it allows the engine to run more smoothly, and a smooth-running engine is good because it allows us to get places, and getting places is good . . . . Likewise, cinema going is good, as friendships are good, because they are good for us, and we are good . . . . But then the thought is that the chain of dependence between values must come to an end somewhere for any of the links to make sense. The chain of dependence must come to an end with something that is valuable in itself, or simpliciter. An infinite chain of dependence between values is implausible.<sup>132</sup>

#### 4.4.2 Stage 2

Stage 2 of the argument says that human beings, or rather, valuers, meet the criteria for being valuable simpliciter. Valuers are such that objects and activities of value are good for them—find their point by being engaged with by them—and such that their value does not depend on their contributing to the good of anything or anyone else. Valuers are valuable

---

<sup>132</sup>See Raz (2001, pp. 145-51). Raz says that his is not a point about infinite regress. What he must mean is that the problem is not *merely* that there is an infinite regress, for infinite regresses are sometimes plausible. But I take his point to be that it is not plausible in this case: an infinite regress undermines value explanation.

in themselves, or simpliciter.<sup>133</sup>

#### 4.4.3 Two notes about Stage 1

There are two important things to notice about Stage 1 of the argument. First, the argument cannot be *merely* that for anything to have value there must be something *non-instrumentally* valuable. That is, the argument cannot be that in order for watering the plant to be valuable, it, the plant, must be, or be related to something that is, valuable for its own sake—say, eating the fruit of the plant which brings pleasure (something non-instrumentally valuable for us). This argument, an argument from instrumental to non-instrumental value by regress, is often taken to be almost truistic.<sup>134</sup> Its key premise is that instrumental value is by definition value that conduces to some positive good. The argument is that there can only be instrumental value provided that what is instrumentally valuable is appropriately related to what is non-instrumentally valuable. The assumption is that of a cause and effect pair, the cause is valuable only insofar as it derives its value from the effect, so that without non-instrumental value there can be no valuable effect. Now, I am not taking a stand on whether this is the right account of instrumental value. It might be denied. One might think that the value of a cause in a cause and effect pair does not depend on the effect's having non-instrumental value, but on its standing in the appropriate kind of relationship to the effect. In that case, it is not that the goodness of watering the plant derives from the non-instrumental goodness of the plant. Rather, watering the plant is a good thing to do because of the relation it bears to the plant—the relation of being suitable or of benefit to it,

---

<sup>133</sup>See Raz (2001, pp. 151-58).

<sup>134</sup>See Zimmerman (2005). The argument is defended by Conee (1982). See also discussions by Beardsley (1965), and Harman (1967).

where the value of that relation is foundational.<sup>135</sup>

But the important point for my purposes is that the argument under consideration is considerably more ambitious. It says that the value of non-instrumental goods, like health, or a pleasant afternoon, or a good life, must be valuable for something, valuers, who are valuable in themselves: valuable whether they are valuable for anything or anyone. The notions of non-instrumental value and value simpliciter are different notions, and we should not assume that the same argument works in both cases. It is worth noting that a slide here is likely to be encouraged by use of the term ‘intrinsic value’, which, as noted above, is ambiguous between valuable for its own sake, and valuable simpliciter. My point is that the argument from instrumental to intrinsic value by regress might be sound in the first sense, and unsound in the second.

Second, it is worth noting that stage 1 of the argument bears comparison with Aristotle’s argument for a final end.<sup>136</sup> Aristotle thinks that there must be some one final thing for the sake of which all our more particular actions are undertaken, for without it they will all be undertaken in vain. His line of thought is nicely brought out by way of a series of questions and answers about any proposed course of action.

Why take a trip to Paris? Because I want to improve my French. Why improve your French? Because I want to read Proust in the original. Why read Proust in the original? Because it is one of the great works of fiction, and much is lost in translation. Why read great works? ...<sup>137</sup>

Aristotle’s thought is that the question and answer series will ultimately terminate with: because it will make my life go well—something about which it does not make sense to

---

<sup>135</sup>I am grateful to discussion on this possibility with Brad Weslake.

<sup>136</sup>See *Nicomachean Ethics*, 1094a 18-22.

<sup>137</sup>I take this way of modeling the Aristotelian argument from Vogt (2010).

ask, and why do that? Aristotle's argument is both a metaphysical and a psychological-cum-motivational argument. As a metaphysical argument, it can be treated on a par with the argument schema under consideration, and my remarks about it (below) can be taken to apply to the Aristotelian argument, equally. The argument tends to be defended in its metaphysical form.<sup>138</sup> But as a psychological / motivational argument it might have independent force and warrant independent consideration. On this version of the argument, when we reflect on our motivations for acting, we see that our motivations for our more particular actions are in the service of a single, fundamental motivation for our lives to go well. Without that fundamental motivation, the thought is, we would not be motivated to do anything—our more particular ends would seem pointless to us.<sup>139</sup> Whatever force the argument has as a proposal about the structure of our motivations, it is important to notice that, so construed, it does not license a conclusion about the nature of the final end, for example, that it is valuable in itself or simpliciter. Psychological / motivational arguments do not license metaphysical conclusions.<sup>140</sup> For we should draw a distinction between, as I would put it, our ways of pursuing or valuing things, and the explanation of the value they have.<sup>141</sup>

---

<sup>138</sup>See Broadie (1991) and Lear (2004).

<sup>139</sup>For a qualified defense of Aristotle's argument taken as a thesis about our 'background motivations', see Vogt (2010).

<sup>140</sup>Indeed, I think it is unlikely that the value of a good life is best explained in terms of its being valuable simpliciter. Rather, I think it is best explained in terms of its being (at a minimum, or most fundamentally) good for the person whose life it is. But I will not make the argument here.

<sup>141</sup>This as a way of formulating a distinction influentially drawn by Korsgaard (1983). In Korsgaard's formulation, there are two distinctions: "One is the distinction between things valued for their own sake and things valued for the sake of something else—between ends and means, or final and instrumental goods. The other is the distinction between things which have their value in themselves and things which derive their value from some other source: intrinsically good things versus extrinsically good things" (p. 170). Korsgaard's treatment has been influential, and it has been refined in subsequent discussions. [See Kagan (1998); Rabinowicz and Rønnow-Rasmussen (2000); and Langton (2007)]. The notions of 'intrinsicality' and 'extrinsicality' at issue in Korsgaard's treatment are not fully standard. It is standard to define intrinsic value as

## 4.5 The Alternatives

So much for the second version of the argument schema. What are we to make of it? A clear way to evaluate the argument is to ask whether there are plausible options besides terminating the regress in value simpliciter. Could the chain of dependence between values come to an end with something that is not itself valuable?—call that option (NV). Or could the dependence go on indefinitely?—call that option (IR). Let me take the options in turn.

Raz (2001) argues against (NV):

If A is good for B which is itself devoid of value, that A is good for B is no reason for anyone to do anything, nor a reason for valuing A in any way at all. It is as if A's value is without value. In other words, it, A, is without value. If B is good, but only inasmuch as it is or can be good for C, then whether the value of A means anything (as explained above) depends on whether C is valuable. If A is watering or spraying a protective spray on B, which is a plant—which is good because it enables B to produce C, its fruit—then the value of watering or spraying A depends on whether the fruit is of any value (assuming that B is not valuable in any other way). If there is nothing good in the fruit, what good is watering it? (p. 147).

The claim here is that for the value of something to be explained in terms of the contribution it makes to something else, the thing to which it contributes must itself be valuable, and so on for each of the links in the explanatory chain. As we might put it, each of the nodes in

---

value on the basis of intrinsic properties, and extrinsic value as value on the basis of extrinsic properties. But Korsgaard seems to have in mind non-relational versus relational explanations of the value of things, where her example of a kind of relational explanation is explaining the value of something in terms of its relation to our attitudes (desires or interests). It is for this reason that I would put the first of her distinctions as a distinction in the explanation of value (while recognising that there is more than one axis along which explanations can be given, and for that reason, more than one explanatory distinction); and the second as a distinction in practical reason concerning how we value things. For Korsgaard, we have reason to keep the distinctions apart because we should allow the possibility of there being something the value of which is to be explained relationally, though it is to be valued as an end (for its own sake). An example would be a work of art the value of which is explained in terms of its capacity to enrich our lives in certain ways, though it is to be valued for its own sake. I agree that cases of this kind show that there is a distinction to be drawn here. It would take further argument to make the case that the distinction is a good one, and that it actually has instances. I offer considerations in favour of a distinction of this kind in Section 4 of 'On the Value of Human Beings'.

the chain of dependence must be good. One might be inclined to object on grounds that there is an ordinary use of ‘good for’ on which ‘x is good for y’ and y is not valuable. For example, we might say that while weapons are good for the victors to win the war, their victory is not a good thing. Or, we might say that while certain conditions are good for the growth of a certain bacteria, the growth of that bacteria is not good. But perhaps this is merely a loose way of speaking. Perhaps we are here simply spelling out a causal relationship, so that having bombs would cause or enable the enemy to win the war, as certain chemical conditions enable growth of the bacteria. The proposal might be that we should reserve the expression ‘good for’ for cases where the objects, activities, states, or events being related are valuable, and to that extent, reason giving. As Raz says of what is instrumentally good for something:

It is often assumed that something is instrumentally good just if there is something for which it is good, however valueless, or even bad, that thing is. But if something is instrumentally valuable it is valuable. It is worth having, or keeping or acquiring, or it has exchange value, etc. But if all that can be said for it is that there is something it is good for, e.g., that it is good for making people suffer, it does not follow that it is good. Indeed, as in this example, it may well be bad (p. 146, fn 28).

This seems right to me.<sup>142</sup> In any event, I am prepared to grant that each of the nodes in the chain of dependence must be good. For my purposes, it would not be a happy result if it turned out that valuers—by hypothesis, the substitution instance for the final node in the chain of dependence—were not valuable. What is wanted here, after all, is an account of the value of valuers.<sup>143</sup>

---

<sup>142</sup>In ‘On the Value of Human Beings’ that what is good for something or someone is also good is a claim I show myself to be committed to, and defend.

<sup>143</sup>Of course, that is to set aside the possibility that human beings are not valuable, a view that likely has adherents, and for interesting reasons. It is my understanding that for the ancients, human beings are not bearers of value. Other things have value for us, but we are not valuable. To that extent, the idea that human

What then of option (IR)? If each of the nodes in the chain of dependence must be valuable, why not take it that they are each valuable for something or someone else, and so on ad infinitum? That is, what precisely is wrong with an infinite chain of dependence here? It is something of a philosophical reflex that infinite regresses are thought problematic, though what precisely is amiss is not always made fully clear. It takes some work to spell it out. In this context, one problem is that if we are to imagine an infinite chain of dependence between values that is not circular but linear, so that the value of a node depends on some further node in the chain not yet appealed to, it follows that there would need to be infinitely many things. And it seems implausible to think that the existence of value depends on there being infinitely many things. That is at least one reason to find an infinite regress problematic in this case.

Consider, then, a further option, option (C), on which the dependence between values has a circular structure. There are at least two ways to conceive of this:

On (C<sub>1</sub>): objects and activities of value are valuable because they are valuable for valuers, and valuers are valuable because we are valuable for *objects or activities of value*.

On (C<sub>2</sub>): objects and activities of value are valuable because they are valuable for valuers, and valuers are valuable because we are valuable for *other valuers*.

(C<sub>1</sub>) explains the value of valuers in terms of our being good for such things as mathematics, surgery, global justice, or a valuable relationship. There are two forms this explanation can

---

beings are valuable is a distinctively modern one, and likely has a religious history. [See Vogt (2009)]. My reason for taking human beings to be valuable rests on a set of assumptions about the relationship between reasons, action, and the world. That set of assumptions has been articulated in the writings of Raz over many years, and I am influenced by his treatment. See for example the papers in Raz (1999). As I would put it in the context of human beings, if we think human beings must be treated in certain ways and not in other ways, it is because we take there to be something about them which licenses that treatment. Call that an evaluative feature. This would take some argument to establish, and there is much more to say on both sides. I merely mean to state my auxiliary reason for setting aside the possibility that human beings are not valuable.



take, corresponding to the two kinds of value for: instrumental and non-instrumental value. First, we might be valuable insofar as we make a positive causal difference to things—we advance a discussion, we innovate a technique, or conduct an important study. Second, we might be good for objects of value by being part or constitutive of them, as when we are part of a valuable relationship, or a member of a valuable community.

(C2) explains the value of valuers in terms of our being good for other valuers. Again, there are two forms this explanation can take. We can be good for others by, say, promoting their good. Or we can be good for others by being part of their good, as Harry is good for Harriet in being her friend. To the extent the friendship is a valuable one, Harry enriches Harriet's life—he is part of what makes her life go well. On this model, I am valuable insofar as I contribute to, or am partly constitutive of, the good of others, and others are valuable insofar as they contribute to, or partly constitute, my good (and the good of still others). We can imagine quite complicated forms of interdependence here, with many different kinds of relationship: that between friends, but also neighbours, teachers and students, colleagues, fellow citizens, and so on.<sup>144</sup>

Both (C1) and (C2) are alternatives to terminating the regress in value simpliciter. Is there reason to prefer one candidate over the other? Of (C1), it is worth noting that the explanation of value in each of the cases discussed—advancing a discussion, pioneering an important study, being part of a valuable relationship—can be pushed further. Advancing a discussion is good because it is good for knowledge, we might say, and knowledge is good because it is good for human beings. Conducting an important study is good because it is good for beneficiaries of the study. Relationships are good because they are part of the good

---

<sup>144</sup>That human beings are dependent social beings, and that the good for human beings is interdependent in a robust way, is central to many ancient views in ethics. John Cooper (2010) and Alasdair MacIntyre (1990) develop these lines of thought in Aristotle. Katja Maria Vogt (2008b) has emphasised this line of thought in the Stoa. Samuel Scheffler (1997) has emphasised the importance of at least close personal relationships in contemporary discussions.

of their participants. If that is right, the explanation ultimately turns on our being good for human beings (or other beings).<sup>145</sup> To that extent, (C1) is best construed as a version of (C2). Of course, there are other ways to conceive of the value of valuable things. Another possibility is to take knowledge, relationships (and so on) to be good simpliciter. On this view, human beings are valuable because they make valuable contributions to values, and those values simply have value. Friendship, say, just is valuable, and the value of human beings derives from their participation in the value of friendship. This option seems to me implausible. If one is to invoke the notion of value simpliciter, why invoke it for values, and not for human beings? To do so is to make the value of human beings derivative on the value of something else in a way that calls for explanation.

What then of (C2)? (C2) seems in many ways plausible, and it already presents an alternative to the claim that we must terminate the regress in value simpliciter. One worry might be that (C2) does not capture the intuition that the value of human beings is non-derivative: such that our value does not depend on our being good for anything or anyone else. Recall that this was a key intuition driving the appeal to value simpliciter. Another worry might be that (C2) seems to miss something important about the value of being good for others. Being another's friend or sister or teacher is of value to them, when it is, but it is also of value to *oneself*. Participating in these and other ways in the lives of others, taking on the responsibilities that are part of those forms of participation, may be part of our *own* good. It might seem arbitrary, then, to pick out our relations to *others* as explaining our value, when we often also bear those relations to *ourselves*.

This suggests a third way of spelling out a relational explanation of human value. (C1) says that we are valuable because we are good for objects of value. (C2) says that we are

---

<sup>145</sup>Compare Raz's (1997, p. 296) argument that there is an asymmetry between valuers and values: values are good because they are good for valuers but, at least in the ordinary case, valuers are not good because they are good for values.

valuable because we contribute to the good of others by, for example, standing in valuable relations with them. A third option, call it option (R), starts with the observation that the closest relation we bear is a relationship to ourselves.

On R: objects and activities of value are valuable because they are valuable for valuers, and valuers are valuable because we are such as to stand in a valuable relation to ourselves: a relation of being good for ourselves.

On this option, the chain of dependence between values comes to an end with a kind of reflexive relation.

Now it need not strike us as immediately obvious what it should mean for valuers to be valuable for ourselves. But supposing the idea can be given content, it is noteworthy that (R) has immediate advantages over (C<sub>2</sub>). (R) would make the value of human beings, if not non-relational, then at least non-derivative on the value of anything else. This is important because, without making recourse to a notion of value simpliciter, it would capture an important intuition to the effect that human beings are valuable irrespective of whether we are good for anyone or anything *else*. There would be to that extent something fundamental, or basic, about this explanation of the value of human beings. (R) need not deny that we may be valuable in other ways—that our value may also be explained in terms of our being good-for others, say. It need claim only that being good for ourselves is primary. Further, if (R) is defensible, we can exploit a view on which there is only one kind of value—value for something or someone. Accordingly, the value of human beings, as befits beings who are after all part of nature, will not be thought different in kind from the value of other kinds of things.

What is more, I think (R) can exploit an account of the basis of human value which is independently plausible. Let me now say something briefly about the basis of our value, and by way of doing that, let me further explicate the proposal.

## 4.6 Being valuable for Ourselves

We can develop option (R), and give content to the notion of being good for ourselves, by drawing on Raz's and Korsgaard's intuition to the effect that human beings are valuable in virtue of being *valuers*. Now there are different ways to understand what is to be a valuer, and Raz and Korsgaard offer differing accounts of it. At a minimum, I take it that to value is to pursue projects, interests, relationships, and self-ideals for their own sake. To value is to have commitments or goals in terms of which we structure our more particular actions. This being so, to be a valuer is to be a certain kind of agent, one who acts in sustained and coherent ways. The relevant question to ask, at this point, is why being a certain kind of agent, one who acts in sustained and coherent ways, makes us valuable? On the proposal we are exploring, being a valuer makes us valuable, not simpliciter, but *for* something or someone—in particular, it makes us valuable for ourselves. What is it to be valuable for ourselves? One possibility is that to be valuable for *ourselves* is to be valuable for our *lives*. On this construal, to value is to pursue goals and commitments the pursuit of which makes us valuable because, the thought would be, it makes our lives go well. Then we must ask about the value of a good life. If we are not to make the value of a good life foundational by making it valuable simpliciter, a good life will need to be valuable because it is valuable for something or someone. Plausibly, a good life is valuable most basically for the person whose life it is. (Most basically because, likely, it will be valuable for others too). As it begins to be explicated, then, proposal (R) comes to have the structure of (R\*):

On (R\*): objects and activities of value are valuable for valuers, and being a valuer is valuable because it makes us capable of living valuable lives, and valuable lives are valuable because they are valuable for the person whose lives they are.

What terminates the chain of dependence on (R\*) is less a reflexive relation than a circle. The direction of value explanation takes us from objects of value to valuers, from valuers to lives, and from lives back to valuers. The circle is non-vicious to the extent that different aspects of the valuer is appealed to. Plausibly, we appeal first to whatever particular cognitive and dispositional features make for valuing, in the first instance, and to the individual being or organism as a whole, in the second. Clearly, I have given the barest sketch of how this alternative might be developed. It remains to give a detailed account of valuing, and it remains to make the case that being a valuer in fact makes our lives go well. I cannot hope to make those cases here, but I make some headway elsewhere.<sup>146</sup>

#### 4.7 Conclusion

The task of the present paper has been to present various alternatives to terminating a chain of dependence between values in value simpliciter. I take there to be at least two plausible alternatives, what I have called options (C2) and (R\*). I have given reason to prefer (R\*) on grounds that it preserves an important desideratum in an account of the value of human beings, namely, that our value is non-derivative on the value of anything else. I have sought to develop (R\*) by drawing on the views of Korsgaard and Raz to the effect that what makes human beings valuable is that we are valuers. If R\* can be suitably explicated and defended we can retain the idea that human beings are valuable in virtue of being valuers at the same time as we take the value of human beings to be continuous with the value of other things in being valuable for something or someone. At least, the possibility of a plausible alternative to terminating the regress in value simpliciter gives us reason to develop an account of the value of human beings without making recourse to the notion of value simpliciter.

---

<sup>146</sup>see 'On the Value of Human Beings'.

## 5 On the Value of Human Beings

### 5.1 Introduction

I begin with a very basic thought. We think human beings should be treated in certain ways and not in others—that we are subjects of ethical concern. We needn't think human beings are unique in this respect. We think there are ways we should and should not treat animals, and alien beings (should any be discovered). And there is a case to be made for ancient ruins, and mountain ranges; for childhood keepsakes, and works of art. Whereas once, at least in some quarters, all of ethical theory was thought to fall out of a doctrine of respect for human beings, this is now widely recognised to be implausible. Still, we can identify a mistake while we see the impulse to make it: what we owe to human beings is a part, a very central part, of ethics.

From what we owe to human beings it is a short, albeit not a trivial, step to the *value* of human beings. We take that step when we ask *what it is* about human beings that makes us subjects of ethical concern. That we 'have an elbow', or 'a vestigial appendage', that we 'walk upright', are so many joking candidates; but they make a serious point.<sup>147</sup> If human beings are such as to merit being treated in certain ways, it is because something about us is good—is valuable.<sup>148</sup> The question of this paper is: what makes human beings valuable, and accordingly, such as to be treated in certain ways and not in others?

The question is ambiguous. When we ask about the value of human beings we may be asking about the feature or features on which our value depends—are we valuable in virtue of being moral agents, or valuers in general; in virtue of being rational, self-conscious,

---

<sup>147</sup>They are the joking candidates of Robert Nozick (1981, p. 451).

<sup>148</sup>Though there are differences in the notions 'good' and 'valuable' in ordinary usage, I will use them interchangeably in what follows.

or something else? Or we may be asking for an explanation of why the feature or features confer value on us—does the feature simply make us valuable—valuable simpliciter? Or does the feature make us valuable because it makes us good-for something or someone? The question of the value of human beings in this way sets the stage for three tasks. One, call it *basis*, is to make out the value-conferring feature of human beings. A second, call it *explanation*, is to give an account of what makes that feature value-conferring. A third, call it *response*, is to show how ethical responses to human beings are responses to the value-conferring feature of us.

This paper forms part of an argument by inference to the best explanation for an account of the value of human beings. In full, the argument aims to establish that the account provides an explanation of the value of human beings, and that this explanation is preferable to the alternatives. Though I will at points remark on the comparative explanatory merits of my account, my focus here is on providing the explanation. That is, in this paper, my central aim is to describe a possible basis of human value, to show that it provides an explanation of human value, and to describe the reasons to which it gives rise. On my account, the *basis* of human value is our capacity to have final ends, that is, as a first approximation, the capacity to pursue interests, projects, self-ideals, and relationships for their own sake. I will propose an account of what it is to have final ends, but it is the novel *explanation* of how this grounds human value that is of distinct philosophical interest. It rests on a plausible conception of the nature of value in general, and has been overlooked in the literature.

The structure of the paper is as follows. I begin, in Section 5.2, with an analysis of what it is to have final ends. Adapting a proposal of Samuel Scheffler (2010), I argue that the capacity to have final ends is constituted by a set of cognitive, emotional, and deliberative dispositions. To have a final end is (i) to believe that the end is (non-instrumentally) valuable, (ii) to be guided by the end in long-range deliberation, (iii) to have a stable

disposition to treat the end as reason-giving in relevant deliberative contexts, and (iv) to be emotionally susceptible to successes and failures in one's pursuit of the end.

I turn, in Section 5.3, to *explanation*. My explanation of the value of human beings depends on the notion of benefit or goodness-for. I argue that having final ends makes human beings valuable because it makes us capable of being good-for something or someone. Who or what are human beings capable of being good-for? I make the case that having final ends makes us capable, in the first place, of being *good-for ourselves*. That is, on my view, the primary explanation of the value of human beings turns on our capacity to be good-for ourselves. Note that there are several axes along which explanatory questions may be asked. They may be metaphysical or more broadly meta-ethical. The explanatory question that interests me here is whether value is a relational or a non-relational property in the sense that is at issue in the distinction between value-for and value simpliciter. I am not here asking meta-ethical questions, for example, about realism and anti-realism. Those questions are off-stage for present purposes, though they are no doubt relevant for a full account of the nature of the value of human beings.

The claim that our value turns on our being good-for ourselves has the air of a paradox. It raises a question about the reasons we have to *respond* to human beings. If the capacity to have final ends makes us valuable because it makes us good-for ourselves, what reason should others have to respond to us? This is an instance of a more general worry, raised for example by Donald H. Regan (2004), about the normative force of goodness-for. In Section 5.4, I argue that the worry rests on a misunderstanding about the theoretical role of good-for. By examining a response to this worry due to Joseph Raz (2004), I show how my account of the value of human beings gives rise to reasons to respond to human beings that are reasons for everyone.

More particularly, in Section 5.5, I show that by taking the basis of human value to



be the capacity to have final ends, we can help ourselves to an account of what we owe to others that has some claim to being Kantian (though it is grounded in a non-Kantian theory of value). On this conception, the value of human beings gives rise to reasons: (i) not to impede or destroy the ability of others to have final ends, and (ii) to further or promote the final ends of others.

The resulting proposal is distinctive in a number of ways. In the literature on the value of human beings, the chief task is often taken to be that of determining the *basis* of human value. The question of how the proposed basis *explains* the value of human beings is often neglected. This is particularly evident in the Kantian literature. There, interpreters see the main question as being whether the basis of human value is the capacity for autonomy, or the capacity for setting ends, or a good will (or something else).<sup>149</sup> But interpreters seldom consider how these proposed bases confer value on their possessors, and what the status of the value is meant to be.<sup>150</sup>

Where the explanatory question is taken up, a traditional view is that human beings are valuable in themselves, or good simpliciter. Something is good simpliciter *iff* its value does not depend on its being good-for anything or anyone. This is sometimes put in the language of absolute, intrinsic, or non-relational, value. I depart significantly from traditional explanations in making no appeal to these value notions.<sup>151</sup> A major motivation for explaining human value in terms of good simpliciter is that explanations in terms of good-for have not been thought forthcoming. The impetus of this paper is to offer a viable explanation of the value of human beings which turns on our being good-for something

---

<sup>149</sup>This is the guiding question in the interpretive disputes between, among others, Christine Korsgaard (1986a), Allen Wood (1998), Paul Guyer (1998b) and Richard Dean (2006).

<sup>150</sup>A notable exception is Oliver Sensen (2009a,b).

<sup>151</sup>That human beings are good simpliciter is the position taken in Korsgaard (1983, 1986a), and Raz (2001).

or someone.<sup>152</sup> This is to employ an explanation which makes our value continuous with the value of other things. My task is to show that the account nevertheless captures what is distinctive about human beings, and what is owed to human beings.

## 5.2 The Basis of Human Value

Consider what we are asking about when we ask about the *basis* of value. Imagine we are in the market for a transistor radio, and ask our friend what makes hers a good model. Or imagine we are standing before a work of art, before one of Constantin Brancusi's *Bird in Space*, and ask ourselves: what makes it valuable? When our friend points to the portability of her radio, and to the ease with which it picks up a signal, she is speaking to what I am calling the basis question. Likewise, when we try to say something about the sleekness of bronze on marble, or the sense we get of flight, and poise, we are pointing to features of Brancusi's sculpture on which its value supervenes—at least in our estimation.

To make claims about the features on which the value of an object *in part* supervenes is to identify features of the object such that, necessarily, any object with those features, together with other features of the object, has the value in question (and might not have it without them). To make claims about the features on which the value of an object *wholly* supervenes is to identify features of the object such that, necessarily, any object with those features has the value in question (and might not have it without them). These features admit of further categorisation. They can be further categorised as intrinsic or extrinsic properties, as evaluative or non-evaluative, and so on. It is not uncontentious what precisely these distinctions come to, and our classifications will not always be straightforward.<sup>153</sup> But

---

<sup>152</sup>Note that I will use the terms 'valuable-for' and 'good-for' interchangeably.

<sup>153</sup>Is it obvious that a property like portability is a non-evaluative property? What if portability is a reason to buy the radio? On some views about the nature of reasons, portability will be thought to *constitute* the

I leave these issues aside in what follows.

My hypothesis is that the value of human beings wholly supervenes on the *capacity to have final ends*. What does that mean? To give content to the proposal, we need an account of what it is to have final ends. As we will see, the account will be important when we turn to the *explanation* of human value. Providing the account is the task of the present section.

The notion of a final end is familiar from Aristotle's theory of action, and my account owes a clear debt to Aristotle.<sup>154</sup> I share the idea, for example, that an end is something we aim to bring about via our actions, be it a product or an activity. I share the idea, too, that our ends form chains of dependence that are hierarchically structured. One undertakes to sharpen one's pencil with a view to sketching out one's argument, and one sketches out one's argument with a view to writing it up coherently, and one writes up arguments coherently with a view to doing decent philosophy, and one does decent philosophy for its own sake. What it means for the ends of sketching, writing and doing philosophy to be hierarchically related is for the ends further downstream, philosophy in this case, to be those for the sake of which the earlier, the upstream, actions are undertaken. Philosophy is in this way a *final* end. As a final end it rationalises more local ends, and gives them direction and point. We can put this by saying that final ends *guide* our long-range deliberations and give us a plan: to write this novel, to be spontaneous, to restore the vintage motorcycle, to specialise in pediatric surgery.

And yet, the notion of final ends which I will employ departs from Aristotle's in important respects. Aristotle is primarily interested in the idea that there is a *most final*

---

reason to buy it, in which case portability is an evaluative property after all. In that case, anything can be an evaluative property, and the difference between the evaluative and the non-evaluative is not entirely useful.

<sup>154</sup>I draw here from Book I of the *Nicomachean Ethics*.

end.<sup>155</sup> Aristotle tells us that the most final end will be ‘complete’—that for the sake of which everything else is undertaken, though it is not undertaken for the sake of anything else; and ‘self-sufficient’—such that nothing could be added to make it better. I do not assume that there is a most final end, and so, *a fortiori*, I do not assume that a final end will be ‘complete’, and ‘self-sufficient’. My notion of final ends is relatively weak. On my view, a final end can be pursued both for its own sake and for the sake of something else (as a vocation is pursued both for its own sake, and for the sake of gainful employment). A stronger notion of final ends may also require: (i) that a final end be an abstract value (for Aristotle it is happiness, but other candidates might be freedom, justice, and rightness); and (ii) that a final end be something an agent would not revoke, or put into question, when pressed. My account of final ends is not committed to these further ideas.<sup>156</sup>

What is more, my primary interest is in *what it is to have a final end*—in the attitudes and dispositions involved in making something part of our life in the way that final ends are part of our lives. This is not Aristotle’s primary concern, and his claims do not take us very far in this direction. For an account of what it is to have a final end we must turn elsewhere.

To have a final end is to relate oneself to values in some way. A natural question, then, is whether to have a final end *just is* to value something. Samuel Scheffler (2010) gives the most systematic account of valuing in the literature. At the heart of Scheffler’s account is the observation that ‘to value’ is too simply analysed as ‘to desire for its own sake’, or ‘to believe valuable’ (or some combination). We can desire something for its own sake without valuing it (an ice-cream say); and we can believe that something is valuable without valuing it (I can believe that fiction writing is valuable without valuing fiction writing myself). Whereas

---

<sup>155</sup>Note that on some interpretations of Aristotle, the most final end is disjunctive, i.e., a list of various possible high-level final ends.

<sup>156</sup>I am grateful to Katja Vogt for urging me to say more about my conception of ‘finality’.

there seems to be no upper limit to the number of things we can desire, and no upper limit to the number of things we can believe valuable, there does seem to be an upper limit to the number of things we can *value* (p. 21). Valuing seems to involve more, and to have a richer structure, than desiring or believing valuable. Scheffler makes the case for the following four conditions. To value X involves at least (p. 29):

- i) A belief that X is good or valuable or worthy,
- ii) A susceptibility to experience a range of context-dependent emotions regarding X,
- iii) A disposition to experience these emotions as being merited or appropriate,
- iv) A disposition to treat certain kinds of X-related considerations as reasons for action in relevant deliberative contexts.

So if I value my sister, I believe that she (and possibly our relationship) is valuable. But more than that, I am susceptible to disappointment if she cannot make the trip from England, and to delight if good news comes her way. I am disposed to find these appropriate ways to feel about her. Moreover, when she receives the good news, I am inclined to celebrate it with her, and when she makes the trip from England, I make plans to do things I know she will enjoy.<sup>157</sup>

Do Scheffler's conditions capture all we want for an account of having final ends?

It is important to note that his concern is to give an analysis of the ordinary notion of

---

<sup>157</sup>I should note that it is not obvious to me that there is a single notion of 'valuing' at issue in philosophical discussions. While the account of Joseph Raz (2001) bears comparison with that of Scheffler—valuing for Raz involves belief that the object valued is valuable, together with a range of attitudes and responses appropriate to the value (p. 155, fn. 41)—Raz envisages different stages of relating to values. We *acknowledge* values when we form beliefs and other attitudes that are consistent with the value they have (p. 161). We *respect* values when we do not destroy, and perhaps make an effort to preserve, them (p. 162). And we *engage* with values when we participate with values more full-bloodedly: "when we listen to music with attention and discrimination, read a novel with understanding, climb rocks using our skill to cope, spend time with friends in ways appropriate to our relationships with them, and so on and so forth" (pp. 162-3). I will remark on an important point of difference between the accounts of Raz and Scheffler below.

‘valuing’. Among the things we can ordinarily be said to value, for Scheffler, are projects and relationships, but also a family photograph, a friend’s sense of humour, and the opinion of a trusted advisor (p. 15 and *passim*).<sup>158</sup> I am prepared to grant that we use the notion of ‘valuing’ to cover this range of phenomena. But the list makes clear that the notion is importantly different from the notion of having final ends. One cannot have a friend’s sense of humour as a final end, nor a family photograph, nor someone’s opinion. These are not, in the relevant sense, *ends*; nor are they capable of giving deliberative structure to a life in the manner of *final* ends.

Still, Scheffler’s conditions might be thought to capture important components of the final ends idea. The first is the condition of believing valuable. Valuing, we might think, involves believing valuable to the extent that it is a rational attitude. An attitude is rational if it is responsive to reasons, and being responsive to reasons involves (at least tacit) belief in some good-making feature of the object responded to. We form intentions on the basis of this belief. We interpret and explain the actions of others by attributing this belief to them. Or so I will assume.<sup>159</sup> This conception of the relation between agents and value is my motivation for building ‘believing valuable’ into the account of what it is to have final ends. On my view, valuing, like having final ends, is a rational attitude, and accordingly, I will assume that it involves believing valuable.<sup>160</sup>

The second is the condition of emotional susceptibility. For Scheffler, if the photograph one values is destroyed, one is disposed to be upset by its destruction; if the friend loses

---

<sup>158</sup>Compare T. M. Scanlon (1998, p. 95): "We value many different kinds of things, including, at least the following: objects and their properties (such as beauty), persons, skills and talents, states of character, actions, accomplishments, activities and pursuits, relationships and ideals".

<sup>159</sup>For an argument to this end, see Chapter 1 of Joseph Raz (1999); Raz (2010); Raz (2004, pp. 271-2).

<sup>160</sup>Scheffler’s motivation for including the condition of believing valuable is less theoretical. Scheffler takes the condition to be part of the ordinary meaning of ‘to value’. In this he draws a contrast with the ordinary notion of ‘to care’, which, he says, does not involve believing valuable.

her sense of humour, one is disposed to be worried about one's friend—even disenchanted with her. It seems right that having final ends will be partly constituted by this kind of susceptibility. And yet, it seems the character of emotional susceptibility in the case of final ends will have a different structure. Recall that final ends guide one's conception of one's life, and one's long-range deliberations. It seems that one will be vulnerable to how and whether one's conception comes to pass, and to what interim successes and failures mean for one's conception. It will be a susceptibility not in the first place to the object one pursues, but to *one's pursuit* of it—not to the fate of mathematics, but to the fate of one's pursuit of mathematics. The vulnerability will be to how the end turns out for one.<sup>161</sup>

Scheffler's fourth condition concerns dispositions to act. For Scheffler, when one values something, one is disposed to treat it as reason-giving, not merely in a single context—once—but in contexts that are relevantly similar.<sup>162</sup> The condition should capture the sense in which someone cannot value for an instant, but only over time, and with a degree of persistence. As it stands, however, the condition is too weak to capture the temporal dimension. For a person can have and then lose a disposition to treat something as reason-giving in relevant contexts. In a burst of enthusiasm, I can be disposed to treat writing as reason-giving in relevant contexts in the future, and lose that disposition as soon as I sit down to write. To value writing, the disposition to treat writing as reason-giving must itself be *stable*. This seems true of having final ends, too. The disposition to treat an end as

---

<sup>161</sup>Scheffler's third condition is being disposed to find one's emotional susceptibility appropriate. I take this condition to be entailed by the belief that the end is valuable, together with the susceptibility to experience emotions regarding one's success or failure in pursuit of the end. Accordingly, I do not list it as a separate condition. The entailment requires the assumption that believing valuable entails (at least tacitly) believing that emotional susceptibility is appropriate; and it requires the assumption that being emotionally susceptible entails (at least tacitly) believing that one is emotionally susceptible.

<sup>162</sup>That valuing, for Scheffler, is in this way multi-contextual is a point of difference from Raz (2001). At least where Raz speaks of *engaging* values, nothing rules out the possibility that one can properly engage with values on a single occasion, as when one goes to the opera, listens with appreciation and understanding, and never goes again. I will come back to this difference in Section 5.3.

reason-giving in relevant contexts must be a stable feature of one's motivations.<sup>163</sup>

In sum, I take having final ends to include, but to go beyond, valuing in Scheffler's sense. On the model of Scheffler's four conditions, I propose that one has a final end *iff* one:

- i) has a belief that the end is valuable,
- ii) is guided by the end in long-range deliberation,
- iii) has a stable disposition to treat the end as reason-giving in relevant deliberative contexts,
- iv) has a susceptibility to experience a range of appropriate emotions regarding the success or failure of one's pursuit of the end.

So if appreciation of vintage motorcycles is a final end for me, then I believe that appreciation of vintage motorcycles is valuable. But more than that, I am disposed to plan my vacation time around the annual expo, or to restore the motorcycle that has been fifty years in the family. In the meantime, I am disposed to read the relevant trade magazines, and to scour trading sites for hard to find parts. Should I fail to find the signature handle bars for the restoration, then I am disposed to feel frustrated. Should the bike be restored to its former stature, and a glory trip taken around the country, I am disposed to take real pleasure in the execution of my plan.

---

<sup>163</sup>This point about motivational stability is emphasised by Harry Frankfurt (1999) in his account of caring. To care about something, for Frankfurt, is to be committed to caring about it: “[t]he question of whether a person cares about something pertains essentially to whether he is *committed* to his desire for it [...] or whether he is willing and prepared to give the desire up and to have it excluded from the order of his preferences” (p. 161). My view differs from Frankfurt's in that having final ends involves (but is not wholly constituted by) a commitment to having the end. And though Frankfurt takes the commitment to feature in the content of the attitude of caring itself, it need not. The commitment could be a fact about the person. For example, it could be a second-order dispositional fact.



Obviously, there is more to say about final ends. Some final ends will have a clear end-point or moment of completion. Others will be more diffuse, with no clear end-point. It seems possible to have a final end without knowing that one has it as an end, or without knowing that the end is final. It may be that one's final ends reveal themselves in one's behavior, rather than in one's avowals and claims.<sup>164</sup> A person may have several final ends, and they may be of varying importance to her. Final ends may be related to one another in various ways (hierarchically, or as parts of wholes, or in some other way); or they may be largely unrelated. Though it is possible for final ends to be largely unrelated, there may be normative pressure to unify them. There are likely other constraints. For example, a person's final ends cannot be manifestly incompatible with one another (as being a long distance runner is incompatible with being a sumo wrestler).<sup>165</sup> I raise these points to indicate some of the further issues and open questions. But I will largely set them aside in what follows.

### 5.3 The Explanation of Human Value

That it is some kind of involvement with values that makes human beings valuable is a common starting point in accounts of human value.<sup>166</sup> I am taking as a hypothesis that the kind of involvement with values at issue is having final ends, and I have given an account of what it is to have final ends. My strategy is to defend the account by showing how it yields a plausible explanation of the value of human beings. It is to this explanation that I now turn.

So far, in speaking of the features on which the value of an object supervenes, I have said nothing about *why* those features make the objects valuable. Why, to return to our earlier

---

<sup>164</sup>I am grateful to Wolfgang Mann for discussion on this point.

<sup>165</sup>I owe this example to Wolfgang Mann.

<sup>166</sup>It is central to the account of human value in Christine Korsgaard (1983, 1986a, 1996d), and Joseph Raz (1997, 2001), and part of the proposal of Robert Nozick (1981).

examples, is being portable such as to make a radio valuable? And why is being made of bronze, or evoking flight, such as to make Brancusi's sculpture valuable? If the radio is easy to carry, we might take it out to the garden to listen to a program while we prune the fruit trees. Being made of bronze, we might say, the statue has a certain luminosity; while evoking flight, the statue expresses certain physical concepts (space, mass, balance, curvature) in an unusual or imaginative way. These are candidate answers to what I am calling the *explanatory* question: what makes a property, or a set of properties, value-conferring? The first thing to note about the explanations just given is that they are partial. We can raise explanatory questions anew of each of them: what about being able to listen to a program while we work in the garden makes the radio valuable? What about being luminous, and what about the expression of physical concepts, makes the sculpture valuable? Perhaps we will say that working to a program in the garden makes the work pass more quickly. Perhaps we think that, in being luminous, the sculpture affords a pleasant visual sensation, and pleasure just is good—good simpliciter. Perhaps we take the expression of physical concepts to enrich our understanding of how the world works, where understanding the world is part of what enriches a life. It is possible for at least some of these explanations to be pushed further. And it is likely that the explanations given are overly schematic. Presumably, the way in which Brancusi's sculptures enrich our understanding of physical concepts is connected with their capacity to bring pleasure, and so on.

These explanations admit of classification, and this classification is of some importance for what follows. It seems we have broadly two kinds of explanation here. We have features of an object which purport to make it valuable in itself—*valuable simpliciter*; and we have features of an object which purport to make it *valuable-for* something or someone. Moreover, there seem to be two ways something can be valuable-for something or someone. It can be valuable in virtue of the causal contribution it makes to something else of value,

call this *instrumental* value; or it can be valuable in virtue of being part or constitutive of something else of value, call this *non-instrumental* value. So if the explanation of the value of having a pleasant visual sensation is that a pleasant visual sensation just is good, then a pleasant sensation has value simpliciter (value whether or not it is has value-for anything). If the explanation of the good of having a radio in the garden is that it makes our work pass more quickly, then having a portable radio is instrumentally valuable-for someone. If the explanation of the good of understanding the workings of things is that it is part of what enriches our lives, then the value of having an understanding of things is non-instrumentally valuable-for someone.

In setting up the explanatory distinction as a distinction between good and good-for, I am following Raz (2001, pp. 146-151, and passim). I take the distinction between good and good-for to be an *instance* of a more general distinction between non-relational and relational value. Something has non-relational value *iff* its value does not depend on its relationship to something or someone. Something has relational value *iff* its value depends on its relationship to something or someone. It is possible for something to have relational value without being good-for something or someone. For example, the value of a good will for a Kantian may be thought to depend on a relationship to valuers, though its value does not depend on its being good-for something or someone. Some commentators, for example, Michael Smith (2003), treat relational value as synonymous with good-for. Though one can stipulate terms as one sees fit, I think this usage elides important differences. My interest here is in the notion of good-for, and though good-for is a form of relational value, it is a specific kind of relational value.<sup>167</sup> Note that there are important further questions about what it is for something to be valuable simpliciter, and what it is for something to be

---

<sup>167</sup>On relational as against absolute value, see Railton (2003, p. 16, and passim).

valuable-for something or someone.<sup>168</sup> But for present purposes, I will not take a stand on them.

### 5.3.1 The value of human beings as bearers of final ends

So much for the theoretical options for explaining value in general. The argument now proceeds as follows. In ‘On Regress Arguments for the Value of Valuers’ I made conceptual space for an account on which the capacity to pursue final ends makes us valuable by making us valuable-for something or someone. I considered an option on which the capacity to pursue final ends makes us valuable-for other human beings, and I considered an option on which it makes us valuable-for ourselves. For reasons I gave there, it is the second option that I wish to develop. So the question I am asking now is: how does the capacity for having final ends make us valuable by making us good-for ourselves? First we need to give content to the notion of being good-for ourselves. By being good-for *ourselves* I mean that we are good-for our *lives*. The proposal I am considering, then, is that the capacity for having final ends makes us valuable by making us good for our lives. In fact, the proposal is weaker than that. It is that the capacity to pursue final ends makes us *capable of being* good-for our lives. In a clear sense the success of this explanation depends on its being the case that the pursuit

---

<sup>168</sup>There are several competing accounts of ‘good-for’ in the literature. Though their accounts differ in important respects, Richard Kraut (2007) and Connie Rosati (2006) both treat X’s being good-for S as a relation of fit or suitability between X and S, something about which there are attitude independent facts of the matter. There is agreement between them that: (i) good-for should not be understood, perspectively, as good from someone’s point of view; (ii) good-for should not be understood as good and occurs in the life of; (iii) good-for need not be relative to particular agents, but can be relative to groups; (iv) X’s being good-for Y depends on facts about the nature of X and facts about the nature of Y; (v) what is good-for human beings does not depend on (though it may be affected by) what human beings want. Rosati gives a further analysis of the relation of fit, so that ‘good-for’ is a second-order relational property of being productive of a set of features which she enumerates. See Rosati (2006, pp. 118-123). On the differences between Kraut and Rosati, see Rosati (2009). Prominent subjectivist accounts of ‘good-for’ are given by Peter Railton (1986), for whom X is good-for S *iff* X is such as to be desired by S were S fully informed and ideally rational; and by Wayne Sumner (1996), for whom, roughly, X is good-for S *iff* S has a favourable attitude toward X, and that attitude is ‘authentic’.

of final ends makes us capable of living a good life. That is, it must be shown how each of the conditions on having a final end analysed in Section 5.2 is a constituent of a good life. I am committed to treating the four conditions as necessary components of a good life, and as the sufficient components contributed by subjects.<sup>169</sup> Let me consider the conditions in turn.

Is it plausible to think that a good life is a life in which a person pursues some project, self-ideal, or commitment which she *believes to be valuable* (condition I)? I think it is. It would be bad-for someone to pursue a project or commitment which she believed to be disvaluable, or of no independent value. If there is a rational connection between pursuing something and believing that it has some value, then pursuing an end which is believed to be disvaluable, or of no independent value, will be irrational or incoherent. While certain aspects of our lives may be marked by irrationality or incoherence without harm, it would be harmful for our pursuit of final ends to be irrational, since final ends play a structuring role in our long-range deliberations.<sup>170</sup>

Is it plausible to think that a good life is a life in which a person is *guided* by a final

---

<sup>169</sup>The qualification “sufficient components contributed by subjects” is meant to allow that fortune also contributes to our lives going well, but that the contributions of fortune are irrelevant to our value.

<sup>170</sup>Compare Harry Frankfurt. For Frankfurt, caring plays the central deliberative role that I envisage for final ends, but Frankfurt maintains that we can appropriately care about something without believing that it has value. In Frankfurt’s analogy, the perspective of the carer on the object of care is akin to the perspective of a God who loves human beings without regard for their antecedent worth (1988, p. 94). We can care, in Frankfurt’s idiom, for no reason, or at least, for no reason independently of our caring. Frankfurt (2004) cites care of one’s children as paradigmatic here: one does not care about one’s children because one recognises that they have value (p. 39 and *passim*). But is this a clear example of caring for no reason? Perhaps the reason is simply that the children are one’s own. (I owe this point to Katja Vogt). Indeed, it is difficult to find a case in which it can truly be said of someone that she cares for no reason. One might venture Scheffler’s (2010, p. 25) example of a Red Sox fan who cares how his team does against the Yankees (without thinking them more worthy of support). But what would we make of the team supporter whose support was wholly without evaluative thoughts and reasons? It seems to me that his support would itself be lacking something. It is striking that fans tend to yoke their support to narratives that bring out the good-making features of the team, or of its victory: that winning would give a key retiring player a chance at the championship that has always eluded him; that this is a team trained up from grass-roots; that the opposition play a cynical kind of game; and so on.

end in her long-range deliberations and plans (condition II)? To have a final end is to be a certain kind of valuer. But consider the difference between what I will call *basic valuing* and *final ends valuing*. To be a basic valuer is to be guided by normative features of the world in the formation of intentions to act, beliefs, emotions, and other attitudes.<sup>171</sup> It is consistent with basic valuing that our responses to normative features of the world be relatively unsystematic; we respond to this and then that, but our responses need not sum to anything. Final ends valuation *involves* basic valuation—it involves responding to this and that normative feature of the world—but it involves having basic valuation summed, cohered, or unified in specifiable ways. The responses of the valuer with final ends are directed over time. Local responses are guided by longer-range objectives, and this gives them a larger structure and point.

That a good life in this way involves deliberative unity, or cohesion, is an old thought. I can do little more than provide initial motivation for it here. It may be helpful to note that it is compatible with my view that a person has as a final end the self-ideal of being spontaneous, or of engaging with the manifold offerings of the world as they present themselves. And it is compatible with my view that final ends like these contribute to a good life. What I deny is that the life of someone who pursues ends without a structure or plan will be well-going. Should the reader find this too strong, I challenge her to substitute a weaker account of the basis of human value that yields the requisite explanation. Raz (2001, pp. 151-2, f.n 33) draws a distinction between something's being good-for someone, and something's making someone good. He says: "It is, other things being equal, good for me to have this ice cream now. It does not follow that it will make my life, let alone me, better to have it". In a similar way, I would urge that to engage with values in the manner of a basic valuer may be good-for her, but it need not make her valuable. To make her valuable

---

<sup>171</sup>See (Raz, 2001, p. 153).

she would need to have, among other things, basic valuation summed, cohered and unified in the manner of final ends valuing.

The third condition is related to the second, and the argument is the same. Not only must more particular ends be guided by longer-term objectives, but a person must have *stable dispositions* to engage with the end (and end-related activities) over time. Without that, his engagement will be superficial, and may not count as engagement at all. And a life without sustained engagement cannot be well-going.

The fourth condition involves appropriate emotional response to how one's final ends turn out. It seems right that appropriate emotional response will be part of what constitutes a good life. A person with the appropriate emotional responses will find meaning and point in what she does; she will pursue her projects with energy. Dissatisfaction in how her ends turn out will lead her to adjust her plans, or to pursue them differently. It is plausible that someone who neither finds satisfaction in her pursuit of final ends, nor dissatisfaction when her ends encounter difficulties, will not pursue those ends in the right way. And a life without the engaged pursuit of final ends will not be well-going.

Of course, it is beyond the scope of this paper to give a full account of a good life. My interest in outlining these connections between the capacity to have final ends and the nature of a good life is twofold. On the one hand, it helps to show what being good-for ourselves amounts to; on the other hand, it shows how the capacity to have final ends makes us good-for ourselves.

What, then, of the value of a good life? On some views, the existence of a good life will be thought to be a better state of affairs (for the universe) than the existence of a bad life; a good life will be thought to have value in itself, or *simpliciter*. But if we think that something is valuable only if it is capable of being good-for something or someone, as I do, we should say that a good life is valuable because it is good-for the person whose life it is.

We are now in a position to appreciate a refinement to the account. It should be clear from the foregoing that I am conceiving of a good life in normative terms. Some people think we can undertake a value-free investigation into the well-going life for human beings. I think this is implausible. I think that an account of a good life will be normative through and through.<sup>172</sup> So, on my view, someone who pursues ethically bad ends will not have a good life. The refinement is that to be capable of having a good life—and so to be valuable—human beings must be capable of pursuing *valuable* final ends. What is more, they must be capable of pursuing valuable final ends that are appropriate to them, given their aptitudes, interests, cultural background, genetic make-up, and so on. There are many valuable things, but not all valuable things are valuable-for all people. Our choices among values are crucial.<sup>173</sup>

Finally, let me make a remark about capacities. It is plausible that a capacity is valuable only if there is a live possibility that it will be exercised.<sup>174</sup> A person who lives in conditions of deprivation may not now be in a position to exercise her capacity to lead a good life. But she will if conditions change, or she moves elsewhere.<sup>175</sup> Matters are different when someone's capacity has no possibility of being exercised, for example, in cases of profound disability, or irreversible coma. It is implausible to account for the value of these people, and our duties towards them, by saying that they are capable of having final ends, and therewith, of leading a good life. For if it makes sense to say that these people have the relevant capacity—and it may not—it is unclear that the capacity is valuable. My account does

---

<sup>172</sup>For a classic defense of this way of reading Aristotle, see M. F. Burnyeat (1980).

<sup>173</sup>On this point, see Connie Rosati (2006), including her discussion of the sense in which what is good-for us is both discovered and invented.

<sup>174</sup>See Raz (2004, p. 290).

<sup>175</sup>It is plausible to think that it is precisely because people living in conditions of deprivation cannot exercise their capacity for a good life that we are obliged to do what we can to change those conditions.



not provide guidance in these cases. They are difficult cases, and there may be normative reasons not to treat them as more clear than they are.<sup>176</sup> For present purposes I set such cases aside as the topic for another paper.

#### 5.4 The Response to Human Value

My proposal that the value of human beings turns on our being capable of being good-for ourselves has the air of a paradox. It can look as if we have an account of the value of each person to him or herself, but not to others. So it can seem as if we need a further story to take us from value in our own case to value in the case of others. This impression is understandable, but it rests on a misunderstanding about the relationship between the basis of human value, the explanation of human value, and the reasons for action.

One way to put the objection is in terms of the viability of the notion of good-for on which it crucially relies. Donald Regan (2004) has recently raised such a challenge for the good-for theorist. It turns on the normative force of good-for.<sup>177</sup> Good-for is normatively inert, according to Regan, because the good-for someone makes no claims on anyone else. In his example, if the good-for Abel is peculiarly Abel's, then why should Cain care? We think there is a deep connection between values and reasons, so if the good is relative to Abel, it would seem to create only Abel-relative reasons. If Regan is right, the consequences for my account of human value would be very bad indeed. We want to be able to say that *everyone* has reason to treat human beings in certain ways and not in other ways. Put otherwise, we want an account of human value to deliver agent-neutral reasons to respond

---

<sup>176</sup>Compare Matthew Liao (2010) who argues that we can give precise, biological criteria for “human moral status”, where that includes anencephalic infants, people in persistent coma, and comparable cases.

<sup>177</sup>Regan's objection takes the form of a dilemma, but for present purposes I pass over the first horn. In my view, it has been adequately rejected by Raz (2004, pp. 273-276). See also Rosati (2008). Regan's objection is a version of G. E. Moore's (1903, p. 99) argument against egoism. On Moore's argument, see Michael Smith (2003).

to human beings. The immediate task is to meet Regan's objection.

Joseph Raz (2004, pp. 273-276) responds to the objection in the following way. Raz allows that *all* good is good-for.<sup>178</sup> That is, as I would put it, Raz allows that good and good-for are mutually entailed: something is good *iff* it is good-for something or someone. Still, there is a question of which, good or good-for, has explanatory priority. As we might put the question: is X good because it is good-for something or someone? Or is X good-for something or someone because it is good? Raz takes the second option: X is good-for something or someone because it is good. Good has explanatory priority, and it is for this reason that what is good, though it is necessarily good-for something or someone, can give rise to reasons for everyone. From the explanatory priority of good over good-for, Raz rejects Regan's challenge. He adduces the following piece of linguistic data. We say: "It would be good for you to read this novel. It's really excellent". (So 'good', or as here 'excellence', has explanatory priority over 'good-for'). Moreover, it is that very quality, being excellent, which makes it good-for others too. It is good-for you for the same reason that it is good-for me: it is an excellent book. (So though values are valuable-for something or someone, they can give rise to reasons for everyone).

While Raz secures the conclusion we are after, he does so at a significant cost. Raz makes an important concession to the Regan style view. Though he allows that all values are value-for something or someone, he denies that good just is good-for.<sup>179</sup> I would defend the stronger claim. I would defend the claim that X is good because it is good-for something or someone. That is to take explanatory priority in the reverse direction. When we explain the value of something we do so by pointing to what it is good-for. A pencil is good because it allows us to make marginalia, a gift is good because it allows us to show gratitude

---

<sup>178</sup>Raz supposes this for the sake of argument here; he denies it elsewhere—for example in Raz (2001).

<sup>179</sup>That good just is good-for is defended by, for example, Kraut (2007).

for hospitality, a trip to the museum is good because it enlivens our studies, and so on. Pencils, gifts, and museums are good because they make a positive difference to something or someone: they are of use or of benefit to something or someone. What are we to make of the linguistic data Raz adduces against this reading of explanatory priority? Raz says we say: “It would be good for you to read this novel. It’s really excellent”. But metaphysical theses cannot be read directly from claims about what we ordinarily say. What matters is not what we *say*, but what makes what we say *true*. Is there a reading of the sentence compatible with the explanatory priority of good-for over good?

On my view, as on Raz’s, the book is excellent—it has value. As with all valuable things, we can say more about that in virtue of which it has value. What makes the book valuable is that it is formally innovative, psychologically insightful and perceptive about the anti-social emotions—about enmity, jealousy, and avarice. These are properties on which the value of the book (at least in part) supervenes; they are the basis of its value. So we might imagine the following conversation:

“It would be good for you to read this book. It’s really excellent”.

“Why? What’s good about it?”.

“Well, it is a novel but it makes use of letters, diary entries and the text of lectures to bring together the world of many characters . . . . The characters each represent one of the vices . . .”

This is to give a further specification of what the value of the book consists in, we might say. And yet, so far, we have said nothing about what makes those properties valuable properties. They could be valuable because they just have value, or because they are valued by human beings, or because they stand in some relation to a stone in the corner of the universe. On the view I am proposing, what explains the value of these properties is that they are (capable of being) good-for readers. (I leave the details of that story to one side here). So what makes

it true that it is an excellent book is that it has properties *xyz*. Properties *xyz* form the basis of the value of the book. But what makes properties *xyz* valuable properties is that they are capable of being good-for readers. This is to give an account of the truth-conditions of Raz's statement that is compatible with treating good-for as explanatorily fundamental.

Now it is from his argument for the explanatory priority of good over good-for that Raz meets Regan's objection. In rejecting Raz's argument have I lost my recourse to his conclusion? I have not. On my view, as on Raz's view, the book has value. And I accept that if something is good-for something or someone, it is good; good and good-for are mutually entailed. What is more, I accept that it is the fact that the book has value (is excellent) that makes it the case that (all things being equal) we should read it. That is, it is the fact that X is good that gives rise to reasons in regard to it, and these are reasons for anyone. I have given a further specification of what the value of the book consists in. It is good in virtue of being formally innovative, psychologically perceptive, and so on. These are properties the book has whether you read it, or whether I read it. What is important is that, from the point of view of reasons for action, it is unimportant how the value of the book is explained: whether it just has value, whether it is related to a stone in the corner of the universe, or whether it is good-for readers. What we should respond to when we respond to books, as to people, is their value. This is Raz's conclusion, and I am perfectly eligible to draw it.

How does all this bear on the account of human value? Our starting point was that human beings are valuable. We asked: what is valuable about human beings—what is the basis of our value? On my view, human beings are valuable in virtue of our capacity for having final ends. But we can ask a further question. We can ask: how does the capacity for having final ends confer value on human beings—what makes the capacity value conferring? On my view, the capacity confers value on us because it makes us capable of being good-for ourselves. I hypothesise that all value is value-for, and that the best way to construe the

value of human beings is by taking us to bear a relation to ourselves—a relation of being good-for ourselves. Now the worry was that in giving this explanation, the account lacks normative force. To adapt Regan's objection, the worry was that if Fitzgerald has value because he is good-for himself, then no-one but Fitzgerald has reason to value him. We are now in a position to see that the worry rests on a confusion about the status of the appeal to good-for. To say that Fitzgerald is capable of being good-for something in virtue of his capacity to have final ends is to explain how the capacity to have final ends makes Fitzgerald valuable. But it remains the case that Fitzgerald is valuable. Moreover, it remains the case that he is valuable in a specifiable way: as capable of having final ends. The solution to Regan's charge is that it is Fitzgerald's value that gives rise to reasons to respond to him in certain ways. More particularly, it is the fact that he is capable of having final ends that makes it the case that we have reasons in regard to him. As we will see, it is the fact that he is capable of having final ends that gives content to the reasons *we* have.<sup>180</sup>

## 5.5 The Responses We Should Have

Throughout this paper I have relied on a three-fold distinction regarding objects of value. There is (i) the *basis* of value, there is (ii) the *explanation* of the value of the basis, and there is (iii) the practical *response* to which the value gives rise. I have argued that the value of human beings gives rise to reasons that are reasons for everyone. I have put this, in Regan's terms, by saying that the value of human beings has normative force. But a full account will show, more particularly, the reasons we have. At least, it will show the basic forms of ethical

---

<sup>180</sup>Frankfurt (1988, p. 93) says we can care about something *because* we believe that the activity of caring itself has value. Caring itself gives rise to reasons to care. So we can properly care about X because we believe that caring about something is good for us (and we believe that we can care about X and not about Y). This is to invoke the good of caring at the wrong theoretical juncture. That caring is good-for us does not give us reason to care, nor does it give other people reason to respond to us as carers. Rather, it explains what is good about caring, and it is the goodness of caring that gives us reasons to care.

consideration that are owed to human beings.

It is a common philosophical assumption that there is a deep connection between values and reasons. This is a metaphysical assumption, but it has epistemological implications. The value of something shows us the reasons we have (and vice versa).<sup>181</sup> Moreover, to the extent that the basis of the value tells us what the value consists in, it is the basis that determines, more particularly, the reasons we have.<sup>182</sup> If the basis of human value is that we have the capacity for having final ends, then our response should be to human beings as bearers of final ends. But to say that we should respond to human beings as bearers of final ends takes only us so far. We need to know, further, the *kind* of response in question.

I maintain that we should respond to human beings as capable of having final ends by, on the one hand, respecting their capacity, and on the other, by promoting their ends. While I have not given a Kantian account of the value of human beings, this is a basically Kantian point. Kant thinks we are to adopt principles of action which further the ends of others. In the *Metaphysics of Morals* he introduces the framework in which we are to do so with metaphors of attraction and repulsion. We are to imagine that our external relations with others are governed by two forces:

---

<sup>181</sup>I have put the point in a way that is neutral about the relative priority of values and reasons. The present point depends only on there being a connection between them. But of course there are larger questions about which has priority. If one thinks that value, and more particularly, value-for, plays an ineliminable theoretical role in explaining the propensity of certain features to give rise to reasons for action, then one is committed to the priority of values over reasons. In that case, one denies that the notion of a ‘reason’ is explanatorily primitive. This is the position taken by Richard Kraut (2010, pp. 452-4), in opposition to T. M. Scanlon (1998, pp. 95-100). In addition to the question about priority, there is a further question about ‘buckpassing’, in the term introduced by Scanlon. For Scanlon, to be good is to have other properties that give us reasons. Goodness is in this way a second-order property, and it is not in itself reason-giving. Understood this way, my account is compatible with buckpassing. On my view, it is fine to say that what gives rise to reasons for action are facts about something’s nature, for example, being such as to have final ends, so that goodness itself plays no direct normative role. But I bypass this question here.

<sup>182</sup>As Raz (2001, pp. 164-5) makes the point, “the value of what has value, and the action it is a reason for, are intrinsically connected”. See also Nozick (1981, p. 452).

The principle of **mutual love** admonishes [human beings] constantly to *come closer* to one another; that of the **respect** they owe one another, to keep themselves *at a distance* from one another (6:449)

As Kant goes on to unpack the metaphors, mutual *love* requires us to adopt principles of active benevolence towards others. Active benevolence expresses itself in a requirement to “make others’ *ends* my own”. That means we are to help others realise their interests and projects, insofar as we do not judge them to be bad for themselves or others. But the requirement to *respect* others means that we are to make the ends of others our own without forcing our ends on them. This calls for non-paternalism on our part.<sup>183</sup>

I find it plausible that a basic form of ethical consideration that we owe to human beings is (i) to respect human beings by not impeding their ability to have final ends, and (ii) to further or promote their ends. There are remaining questions about precisely what behaviours are compatible with (i) and (ii). For Kant, (i) involves the avoidance of contempt, arrogance, and ridicule, where (ii) involves cultivating generosity and sympathetic feeling. Clearly there is more to say. There are questions about whether these further behaviours can be *derived* from the principle to not impede the ability of people to have final ends, and to further their ends.<sup>184</sup> There are also questions about the status of the principle. For Kant and Kantians the principle is categorically binding. While I take the reasons we have to respond to human beings to hold independently of anyone’s desires, since I take the explanation of the reason-giving force of having final ends to turn on its being good-for something or someone, I deny that the principle is categorically binding. It remains to be made out how it compares with other principles or reasons. Likely, the stringency of

---

<sup>183</sup>Onora O’Neill (1985, pp. 264-66) develops this point about non-paternalism. On the tensions between respect and love, but the need to practice them in concert, see Vogt (2008a).

<sup>184</sup>An influential criticism of Kantian attempts to derive concrete principles from the Formula of Humanity is made by Frankena (1986). O’Neill (1985) defends Kant. Raz (2004) denies that derivation is possible, because he takes the basic principle of respect for human beings to take specific cultural forms.

principles or reasons will depend on the character of the value in question—on the features which make it valuable. But these are further matters.

## 5.6 Conclusion

My aim in this paper has been to show how an assumed basis of human value generates an explanation of the value of human beings, and basic forms of ethical response to human beings. I hope to have shown that the resulting account is plausible and attractive. In fact, I believe my account has distinct advantages over accounts which invoke notions of absolute, or non-relational value—what I have called value in itself, or value simpliciter. There are at least three motivations for giving an explanation in those terms. The first stems from an argument from the nature of value, an argument I consider and reject elsewhere.<sup>185</sup> The second stems from the desirability of making the explanation of human beings non-derivative on the value of something else. I have shown that to explain the value of human beings in terms of our being good-for ourselves is to make the value of human beings non-derivative in the relevant sense. The third motivation stems from the would-be distinctiveness of the value of human beings. To explain the value of human beings by appeal to the notion of value simpliciter makes the value of human beings metaphysically distinctive: unlike the explanation of the value of anything else. But I believe the explanation of the value of human beings is *continuous* with the explanation of the value of other kinds of things: radios, works of art, and friendship. So on my view, there is *nothing* metaphysically distinctive about the value of human beings. And yet, we should still think human beings are distinctively valuable. The distinctiveness of human beings, and of what is owed to human beings, is plausibly captured by the distinctiveness of the properties on which our

---

<sup>185</sup>In 'On the Non-Relational Value of Human Beings by Regress'.



value supervenes, and by the distinctiveness of the responses which they demand. We are distinctive in virtue of having final ends, and it is our having final ends that gives rise to reasons to treat us in certain ways, and not others.<sup>186</sup> If that is right, then none of these motivations provide reason to prefer an account in terms of value in itself over an account in terms of value-for something or someone.<sup>187</sup>

---

<sup>186</sup>It is noteworthy that whereas Raz (2001) argues that the nature of the value of human beings is metaphysically distinctive in being non-relational, non-relationality plays no role in generating reasons to respond to human beings in his account. It is the basis of the value—for Raz, that human beings are basic valuers—that does the work.

<sup>187</sup>Many people have provided helpful feedback on this paper. I would like to thank Guy Fletcher, Ben Bramble, Matthew Noah Smith, and Wolfgang Mann for helpful feedback on earlier drafts, and for probing questions. I am very grateful to Brad Weslake for incisive editorial suggestions, and ongoing philosophical exchange. My debt to Joseph Raz is large, and will be obvious. I owe great thanks to Katja Vogt with whom I have worked out my views on human value in conversations spanning some years.

## References

- Aristotle. 1987. *A New Aristotle Reader*, edited by J. L. Ackrill. Princeton, NJ: Princeton University Press.
- Annette C. Baier. 1994. *Moral Prejudices: Essays on Ethics*, Cambridge MA: Harvard University Press.
- Marcia W. Baron. 1995. *Kantian Ethics Almost Without Apology*, Ithaca and London: Cornell University Press.
- Anne Margaret Baxley. 2010. *Kant's Theory of Virtue: The Value of Autocracy*, Cambridge: Cambridge University Press.
- Monroe C. Beardsley. 1965. "Intrinsic Value", in *Philosophy and Phenomenological Research*, Vol. 26, pp. 1–17.
- Lewis White Beck. 1960. *A Commentary on Kant's Critique of Practical Reason*, Chicago: The University of Chicago Press.
- Monika Betzler. 2008. *Kant's Ethics of Virtue*, edited by Monika Betzler. Berlin and New York: Walter de Gruyter.
- Heiner Bielefeldt. 2001. "Hegelianizing Kant? Allen W. Wood's Interpretation of Kant's Ethical Thought", in *Canadian Journal of Philosophy*, Vol. 31, No. 3, Sept. 2001, pp. 445–51.
- Sarah Broadie. 1991. *Ethics with Aristotle*, Oxford and New York: Oxford University Press.
- M. F. Burnyeat. 1980. "Aristotle on Learning to be Good", in *Essays on Aristotle's Ethics*, edited by Amelie Oksenberg Rorty, University of California Press, Berkeley, pp. 69–92.
- Earl Conee. 1982. "Instrumental Value without Intrinsic Value?", in *Philosophia*, Vol. 11, pp. 345–59.
- John Cooper. 2010. "Political Community and the Highest Good", in *Being, Nature and Life: Essays in Honor of Allan Gotthelf*, edited by James G. Lennox and Robert Bolton, Cambridge University Press, Cambridge, pp. 212–264.
- Daniel Dahlstrom. 1993. "Review of *The Metaphysics of Morals*", in *The Review of Metaphysics*, Vol. 46, No. 4, June 1993, pp. 850–853.

- Stephen Darwall. 2008. "Kant on Respect, Dignity and the Duty of Respect", in *Kant's Ethics of Virtue*, edited by Monika Betzler, De Gruyter, Berlin, pp. 175–200.
- Richard Dean. 2006. *The Value of Humanity in Kant's Moral Theory*, Oxford: Oxford University Press.
- David Enoch. 2006. "Agency Shmagency: Why Normativity Won't Come from What is Constitutive of Action", in *The Philosophical Review*, Vol. 115, No. 2, Apr. 2006, pp. 169–198.
- John Martin Fischer and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.
- William J. Fitzpatrick. 2005. "The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism and the Nature of Normativity", in *Ethics*, Vol. 115, July 2005, pp. 651–691.
- Philipa Foot. 1985. "Utilitarianism and the Virtues", in *Mind*, Vol. 93, No. 374, Apr. 1985, pp. 196–209.
- W. K. Frankena. 1986. "The Ethics of Respect for Persons", in *Philosophical Topics*, Vol. 14, No. 2, pp. 149–67.
- Harry G. Frankfurt. 1971. "Freedom of the Will and the Concept of a Person", in *The Journal of Philosophy*, Vol. 68, No. 1, Jan. 1971, pp. 5–20.
- . 1999. *Necessity, Volition and Love*, Cambridge: Cambridge University Press.
- . 2004. *The Reasons of Love*, Princeton and Oxford: Princeton University Press.
- Paul W. Franks. 2005. *All or Nothing: Systematicity, Transcendental Arguments, and Skepticism in German Idealism*, Cambridge MA: Harvard University Press.
- Patrick Frierson. 2003. *Freedom and Anthropology in Kant's Moral Philosophy*, Cambridge: Cambridge University Press.
- Peter Geach. 1956. "Good and Evil", in *Analysis*, Vol. 17, No. 2, Dec. 1956, pp. 33–42.
- Mary J. Gregor. 1963. *Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik der Sitten*, New York: Basil Blackwell.
- Paul Guyer. 1998a. "Introduction", in *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*, edited by Paul Guyer, Rowman and Littlefield, Lanham, pp. xi–xiv.

- Paul Guyer. 1998b. "The Value of Reason and the Value of Freedom", in *Ethics*, Vol. 109, No. 1, Oct. 1998, pp. 22–35.
- Gilbert H. Harman. 1967. "Toward a Theory of Intrinsic Value", in *The Journal of Philosophy*, Vol. 64, No. 23, Dec. 1967, pp. 792–804.
- Barbara Herman. 1983. "Integrity and Impartiality", in *Monist*, Vol. 66, pp. 233–50. Reprinted in Herman (1993b, pp. 23-44).
- . 1985. "The Practice of Moral Judgment", in *Journal of Philosophy*, Vol. 82, pp. 414–436. Reprinted in Herman (1993b, pp. 73-93).
- . 1993a. "Leaving Deontology Behind", in *The Practice of Moral Judgment*, Harvard University Press, Cambridge MA, pp. 208–240.
- . 1993b. *The Practice of Moral Judgment*, Cambridge MA: Harvard University Press.
- . 1996. "Making Room for Character", in *Aristotle, Kant, and the Stoics*, edited by Stephen Engstrom and Jennifer Whiting, Cambridge University Press, Cambridge, pp. 36–60.
- . 1997. "A Cosmopolitan Kingdom of Ends", in *Reclaiming the History of Ethics: Essays for John Rawls*, edited by Barbara Herman Andrews Reath and Christine Korsgaard, Cambridge University Press, Cambridge, pp. 187–213. Reprinted in Herman (2007, pp. 51-78).
- . 1998. "Training To Autonomy: Kant and the Question of Moral Education", in *Philosophers on Education: New Historical Perspectives*, edited by Amelie Oksenberg Rorty, Routledge, London, pp. 255–272. Reprinted in Herman (2007, pp. 130-153).
- . 2002. "Bootstrapping", in *Contours of Agency: Essays on Themes from Harry Frankfurt*, edited by Sarah Buss and Lee Overton, MIT Press, Cambridge MA, pp. 253–274. Reprinted in Herman (2007, pp. 154-175).
- . 2007. *Moral Literacy*, Cambridge MA: Harvard University Press.
- Pamela Hieronymi. 2001. "Articulating an Uncompromising Forgiveness", in *Philosophy and Phenomenological Research*, Vol. 62, No. 3, May 2001, pp. 529–555.
- Rosalind Hursthouse. 1999. *On Virtue Ethics*, Oxford and New York: Oxford University Press.

- Brian Jacobs and Patrick Kain. 2003. *Essays on Kant's Anthropology*, edited by Brian Jacobs and Patrick Kain. Cambridge: Cambridge University Press.
- Robert N. Johnson. 2007. "Value and Autonomy in Kantian Ethics", in *Oxford Studies in Metaethics Volume II*, edited by Russ Shafer-Landau, Oxford University Press, Oxford, pp. 133–48.
- Shelly Kagan. 1998. "Rethinking Intrinsic Value", in *The Journal of Ethics*, Vol. 2, pp. 277–297.
- Immanuel Kant. 1996a. *Practical Philosophy*, edited by Mary J. Gregor with a General Introduction by Allen Wood. Cambridge: Cambridge University Press.
- . 1996b. *The Metaphysics of Morals*, edited by Mary Gregor. Cambridge: Cambridge University Press.
- . 1997a. *Lectures on Ethics*, edited by Peter Heath and J. B. Schneewind. Cambridge University Press.
- . 1997b. *The Critique of Practical Reason*, edited by Mary Gregor. Cambridge: Cambridge University Press.
- . 1997c. *The Groundwork of the Metaphysics of Morals*, edited by Mary Gregor. Cambridge Texts in the History of Philosophy. Cambridge: Cambridge University Press.
- . 1998. *Critique of Pure Reason*, edited by Paul Guyer and Allen Wood. Cambridge University Press.
- . 2006. *Anthropology from a Pragmatic Point of View*, edited by Robert B. Louden. Cambridge: Cambridge University Press.
- Christine Korsgaard. 1983. "Two Distinctions in Goodness", in *The Philosophical Review*, Vol. 92, No. 2, Apr. 1983, pp. 169–95.
- . 1986a. "Kant's Formula of Humanity", in *Kant-Studien*, Vol. 77, pp. 183–202.
- . 1986b. "The Right To Lie: Kant on Dealing with Evil", in *Philosophy and Public Affairs*, Vol. 15, No. 4, pp. 325–349. Reprinted in Korsgaard (1996a).
- . 1992. "Creating the Kingdom of Ends: Reciprocity and Responsibility in Human Relations", in *Philosophical Perspectives*, Vol. 6, pp. 305–332. Reprinted in Korsgaard (1996a, pp. 188–221).

- Christine Korsgaard. 1996a. *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press.
- . 1996b. “Kant’s Analysis of Obligation: The Argument of Groundwork I”, in *Creating the Kingdom of Ends*, Cambridge University Press, Cambridge, pp. 43–76.
- . 1996c. “Morality as Freedom”, in *Creating the Kingdom of Ends*, Cambridge University Press, Cambridge, pp. 159–187.
- . 1996d. *The Sources of Normativity*, Cambridge: Cambridge University Press.
- . 1997. “Introduction”, in *The Groundwork of the Metaphysics of Morals*, edited by Immanuel Kant, Cambridge Texts in the History of Philosophy, Cambridge University Press, Cambridge, pp. vii–xxx.
- . 2009. *Self-Constitution: Agency, Identity and Integrity*, Cambridge: Cambridge University Press.
- Richard Kraut. 2007. *What is Good and Why*, Cambridge MA and London: Harvard University Press.
- . 2010. “What is Intrinsic Goodness?”, in *Classical Philology*, Vol. 105, No. 4, Special Issue: Beauty, Harmony and the Good 2010, pp. 450–462.
- . 2011. *Against Absolute Goodness*, Oxford and New York: Oxford University Press.
- Manfred Kuehn. 2006. “Introduction”, in *Anthropology from a Pragmatic Point of View*, edited by Robert B. Loudon, Cambridge University Press, Cambridge, pp. vii–xxix.
- Rae Langton. 2007. “Objective and Unconditioned Value”, in *Philosophical Review*, Vol. 116, No. 2, pp. 157–185.
- Gabriel Richardson Lear. 2004. *Happy Lives and the Highest Good: An Essay on Aristotle’s Nicomachean Ethics*, Princeton, NJ: Princeton University Press.
- S. Matthew Liao. 2010. “The Basis of Human Moral Status”, in *Journal of Moral Philosophy*, Vol. 7, No. 2, pp. 159–179.
- Robert B. Loudon. 1986. “Kant’s Virtue Ethics”, in *Philosophy*, Vol. 61, No. 238, Oct. 1986, pp. 473–489.

- Robert B. Loudon. 2000. *Kant's Impure Ethics: From Rational Beings to Human Beings*, Oxford and New York: Oxford University Press.
- Alasdair C. MacIntyre. 1990. *Dependent Rational Animals: Why Human Beings Need the Virtues*, The Paul Carus Lectures. Illinois: Open Court.
- Adrienne M. Martin. 2006. "How to Argue for the Value of Humanity", in *Pacific Philosophical Quarterly*, Vol. 87, pp. 96–125.
- John McDowell. 1998. *Mind, Value and Reality*, Cambridge MA: Harvard University Press.
- G. E. Moore. 1903. *Principia Ethica*, Cambridge: Cambridge University Press.
- Iris Murdoch. 1970. *The Sovereignty of Good*, London: Routledge and Kegan Paul.
- Susan Neiman. 1994. *The Unity of Reason*, New York and Oxford: Oxford University Press.
- Robert Nozick. 1981. *Philosophical Explanations*, Cambridge MA: Harvard University Press.
- Martha Nussbaum. 1999. "Virtue Ethics: A Misleading Category?", in *The Journal of Ethics*, Vol. 3, No. 3, pp. 163–201.
- Emer O'Hagan. 2004. "Practical Identity and the Constitution of Agency", in *The Journal of Value Inquiry*, Vol. 38, pp. 49–59.
- Onora O'Neill. 1984. "Kant After Virtue", in *Inquiry*, Vol. 26, pp. 387–405. Reprinted in O'Neill (1989), pp. 145–162).
- . 1985. "Between Consenting Adults", in *Philosophy and Public Affairs*, Vol. 14, No. 3, pp. 252–277.
- . 1989. *Constructions of Reason*, Cambridge: Cambridge University Press.
- David Pugmire. 1994. "Real Emotion", in *Philosophy and Phenomenological Research*, Vol. 54, No. 1, Mar. 1994, pp. 105–122.
- Wlodek Rabinowicz and Toni Rønnow-Rasmussen. 2000. "A Distinction in Value: Intrinsic and For Its Own Sake", in *Proceedings of the Aristotelian Society, New Series*, Vol. 100, pp. 33–51.
- Peter Railton. 1986. "Moral Realism", in *The Philosophical Review*, Vol. 95, No. 2, Apr. 1986. Reprinted in Railton (2003), pp. 3–42).

- Peter Railton. 2003. *Facts, Values and Norms: Essays Towards a Morality of Consequence*, Cambridge: Cambridge University Press.
- Joseph Raz. 1986. *The Morality of Freedom*, Oxford: Clarendon Press.
- . 1997. “The Amoralist”, in *Ethics and Practical Reasoning*, edited by Garrett Cullity and Berys Gaut, Oxford University Press, Oxford, pp. 369–398. Reprinted in Raz (1999, pp. 273–302).
- . 1999. *Engaging Reason: On the Theory of Value and Action*, Oxford and New York: Oxford University Press.
- . 2001. *Value, Respect and Attachment*, Cambridge: Cambridge University Press.
- . 2003. *The Practice of Value*, edited by R. Jay Wallace. With comments by Christine M. Korsgaard, Robert Pippin and Bernard Williams. Oxford University Press.
- . 2004. “The Role of Well-being”, in *Philosophical Perspectives*, Vol. 18, pp. 269–294.
- . 2010. “On the Guise of the Good”, in *Desire, Practical Reason, and the Good*, edited by Sergio Tenenbaum, Oxford University Press, Oxford.
- Donald H. Regan. 2002. “The Value of Rational Nature”, in *Ethics*, Vol. 112, No. 2, Jan. 2002, pp. 267–291.
- . 2004. “Why am I my Brother’s Keeper?”, in *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, Oxford University Press, Oxford, pp. 202–230.
- Amelie Oksenberg Rorty. 1980. *Essays on Aristotle’s Ethics*, edited by Amelie Oksenberg Rorty. Berkeley: University of California Press.
- Connie Rosati. 2006. “Personal Good”, in *Metaethics After Moore*, edited by Mark Timmons Terry Horgan, Oxford University Press, Oxford, pp. 107–131.
- . 2008. “Objectivism and Relational Good”, in *Social Philosophy and Policy Foundation*, Vol. 25, pp. 314–349.
- . 2009. “Relational Good and the Multiplicity Problem”, in *Philosophical Issues*, Vol. 19, Metaethics 2009, pp. 205–234.



- W. D. Ross. 2002. *The Right and the Good*, edited by Philip Stratton-Lake. Oxford and New York: Oxford University Press.
- T. M. Scanlon. 1986. "The Significance of Choice", in *The Tanner Lectures on Human Values*, edited by Sterling M. McMurrin. Vol. 8, University of Utah Press, Salt Lake City, pp. 149–216.
- . 1998. *What We Owe to Each Other*, Cambridge MA: Harvard University Press.
- . 2003. "Metaphysics and Morals", in *Proceedings and Address of the American Philosophical Association*, Vol. 77, No. 2, Nov. 2003, pp. 7–22.
- Samuel Scheffler. 1994. *Human Morality*, New York: Oxford University Press.
- . 1997. "Relationships and Responsibilities", in *Philosophy and Public Affairs*, Vol. 26, No. 3, Summer 1997, pp. 189–209.
- . 2010. "Valuing", in *Equality and Tradition: Questions of Value in Moral and Political Theory*, pp. 15–40.
- J. B. Schneewind. 1998. "Korsgaard and the Unconditional in Morality", in *Ethics*, Vol. 109, No. 1, Oct. 1998, pp. 36–48.
- Oliver Sensen. 2009a. "Dignity and the Formula of Humanity", in *Kant's Groundwork: A Critical Guide*, edited by Jens Timmermann, Cambridge University Press, Cambridge. Chapter 6, pp. 102–118.
- . 2009b. "Kant's Conception of Inner Value", in *European Journal of Philosophy*, pp. 1–19.
- Nancy Sherman. 1997. *Making a Necessity of Virtue: Aristotle and Kant on Virtue*, Cambridge: Cambridge University Press.
- Michael Smith. 2003. "Neutral and Relative Value After Moore", in *Ethics*, Vol. 113, Mar. 2003, pp. 576–598.
- Lawrence Stern. 1974. "Freedom, Blame and Moral Community", in *Journal of Philosophy*, Vol. 71, No. 3, Feb. 1974, pp. 72–84.
- Michael Stocker. 1976. "The Schizophrenia of Modern Ethical Theories", in *The Journal of Philosophy*, Vol. 73, No. 14, Aug. 1976, pp. 453–466.

- P. F. Strawson. 1962. "Freedom and Resentment", in *Proceedings of the British Academy*, Vol. 48, pp. 1–25.
- Roger J. Sullivan. 1996. "Introduction", in *The Metaphysics of Morals*, edited by Mary Gregor, Cambridge University Press, Cambridge, pp. vii–xxvi.
- Wayne Sumner. 1996. *Welfare, Happiness, and Ethics*, Oxford: Clarendon Press.
- David Sussman. 2003. "The Authority of Humanity", in *Ethics*, Vol. 113, No. 2, Jan. 2003, pp. 350–366.
- Judith Jarvis Thomson. 2001. *Goodness and Advice*, edited by Any Gutmann. The Tanner Lectures on Human Values. Princeton, NJ: Princeton University Press.
- . 2008. *Normativity*, The Paul Carus Lectures. Illinois: Carus Publishing Company.
- Jens Timmerman. 2007. *Kant's Groundwork of the Metaphysics of Morals: A Commentary*, Cambridge: Cambridge University Press.
- Jens Timmermann. 2006. "Value Without Regress: Kant's 'Formula of Humanity' Revisited", in *European Journal of Philosophy*, Vol. 14, No. 1, pp. 69–93.
- David Velleman. 1999. "Love as a Moral Emotion", in *Ethics*, Vol. 109, pp. 338–74.
- . 2006. "A Brief Introduction to Kantian Ethics", in *Self to Self*, Cambridge University Press, Cambridge.
- Kadri Vihvelin. 2004. "Free Will Demystified: A Dispositional Account", in *Philosophical Topics*, Vol. 32, No. 1 & 2, pp. 427–50.
- Katja Maria Vogt. 2008a. "Duties to Others: Demands and Limits", in *Kant's Ethics of Virtue*, edited by Monika Betzler, Walter de Gruyter, Berlin and New York, pp. 219–243.
- . 2008b. *Law, Reason and the Cosmic City: Political Philosophy in the Early Stoa*, Oxford and New York: Oxford University Press.
- . 2009. "Do Human Beings Have Non-Relative Value?", in *forthcoming*.
- . 2010. "Desiring the Good: A Socratic Reading of Aristotle", in *Forthcoming*.

- Gary Watson. 1990. "On the Primacy of Character", in *Identity, Character and Morality*, edited by Owen Flanagan and Amelie Oksenberg Rorty, MIT Press, Cambridge MA, pp. 449–469.
- David Wiggins. 1980. "Deliberation and Practical Reason", in *Essays on Aristotle's Ethics*, edited by Amelie Oksenberg Rorty, University of California Press, Berkeley, pp. 221–240.
- Bernard Williams. 1976. "Persons Character and Morality", in *The Identities of Persons*, edited by Amelie Oksenberg Rorty, University of California Press, Berkeley and Los Angeles, pp. 197–216. Reprinted in Williams (1981, pp. 1-19).
- . 1981. *Moral Luck*, Cambridge: Cambridge University Press.
- Susan Wolf. 2011. "Good-For Nothings".
- Allen Wood. 1998. "Humanity As End In Itself", in *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*, edited by Paul Guyer, Rowman and Littlefield, Boston.
- Allen W. Wood. 1991. "Unsociable Sociability: the Anthropological Basis of Kantian Ethics", in *Philosophical Topics*, Vol. 19, No. 1, pp. 325–351.
- . 1996. "General Introduction", in *Practical Philosophy*, edited by Mary J. Gregor, Cambridge University Press, Cambridge, pp. xiii–xxxiii.
- . 1999. *Kant's Ethical Thought*, Cambridge: Cambridge University Press.
- Michael J. Zimmerman. 2005. "Introduction", in *Recent Work on Intrinsic Value*, edited by Toni Rønnow-Rasmussen and Michael J. Zimmerman. Vol. 17, Library of Ethics and Applied Philosophy, Springer, Dordrecht, pp. xiii–xxxv.