

Humanity as an Idea, as an Ideal, and as an End in Itself

RICHARD DEAN

California State University Los Angeles

Email: rdean@calstatela.edu

Abstract

Kant emphasizes that moral philosophy must be divided into two parts, a ‘purely rational’ metaphysics of morals, and an empirical application to individuals, which Kant calls ‘moral anthropology’. But Kant gives humanity (*die Menschheit*) a prominent role even in the purely rational part of ethics – for example, one formulation of the categorical imperative is a demand to treat humanity as an end in itself. This paper argues that the only concepts of humanity suited to play such a role are the rational idea of humanity, and the rational ideal derived from this idea, which Kant discusses in *Critique of Practical Reason* and other texts.

Keywords: end in itself, humanity, idea, ideal, Kant, metaphysics of morals

Introduction

Kant’s emphatic division of moral philosophy into a purely rational ‘metaphysics of morals’ and a more empirically influenced ‘practical anthropology’ or ‘moral anthropology’ poses significant interpretative challenges.¹ One problematic area is the role that humanity can play in a rationally derived metaphysics of morals, since human beings are not purely rational and their characteristics seem to be known partly through empirical observation. Difficult questions arise even at the level of basic moral principles, since Kant says that the categorical imperative is meant to be a supreme principle of morality that is ‘grounded on pure reason alone, independently of all experience’ (Kant 2002: 210; G 4: 409), and yet gives humanity a central role in one version of the categorical imperative, saying that humanity (*die Menschheit*) in oneself and others must be treated as an end in itself (Kant 2002: 230; G 4: 429).

The concept of ‘humanity’ in this ‘humanity formulation’ of the categorical imperative presumably must be accessible through reason alone, if Kant is to maintain a partition between purely rational morality and empirical ‘moral anthropology’. Furthermore, even when Kant turns to the application of basic, purely rational, moral principles to human beings, he at least sometimes maintains that this resulting ‘metaphysics of morals’ must be ‘scrupulously cleansed of everything empirical’ (Kant 2002: 190; G 4: 388). How humanity can be the central object of concern in a rational, *a priori* moral principle, or in a system of moral duties that applies such principles to humans through reason alone, without taking their empirically known features into account, is not obvious.

But I think significant headway can be made on this problem by noting Kant’s technical account, in various writings, of what he calls an ‘idea’ and a corresponding ‘ideal’ of humanity. Within Kant’s overall philosophical system, the idea and the ideal of humanity provide purely rational concepts of humanity suitable for employment at the level of basic principles and of a resulting rational system of more specific duties. This use of Kant’s technical concepts also draws connections between his moral philosophy and other aspects of his Critical philosophy, and sheds light on some otherwise obscure passages in his ethical writings.

1. The Idea of ‘Humanity’ as an End in Itself

Putting aside for now the interpretative issue of the extent to which an overall system of moral duties, or a metaphysics of morals, is meant to be accessible through reason alone, Kant is quite unambiguous in claiming that at least the basic moral principle, or categorical imperative, that underlies all specific moral duties must be known independently of all experience. Given this claim, the central role Kant assigns to humanity in one version of the categorical imperative, which says that humanity in oneself and others must be treated as an end in itself, poses an obvious interpretative puzzle.

Kant has said that basic moral principles must be ‘found completely *a priori* and free from empirical elements in concepts of pure reason’, and that these principles must not be sought for ‘in our knowledge of human nature (which we can get only from experience)’ (Kant 2002: 211; G 4: 410). He adds that the type of end that can play a role in morality must be ‘given by reason alone’ so it can be ‘equally valid for

all rational beings' (Kant 2002: 228; G 4: 427). He even explicitly says that the principle of treating humanity as an end in itself 'is not borrowed from experience' (Kant 2002: 231; G 4: 431). Excluding empirical knowledge of humankind and individual humans from any role in the humanity formulation of the categorical imperative apparently requires some concept of humanity which can be delivered through pure reason, and which can serve as an end in itself.

Kant does describe such a concept of humanity, derived from reason rather than empirical observation, namely what he calls an 'idea' of humanity.² Kant's most explicit discussions of the rational idea of humanity are in *Critique of Pure Reason*, but he seems to employ, and make recurring reference to, this idea of humanity throughout his central writings on moral philosophy.

In *Critique of Pure Reason*, Kant offers a technical definition of an 'idea' as 'a necessary concept of reason', and says 'no congruent object can be given in the senses' (Kant 1998a: 402; CPR A327). Although Kant devotes much more space in *Critique of Pure Reason* to discussing concepts of the understanding, he repeatedly acknowledges that we also have these 'pure concepts of reason', or 'transcendental ideas', which are 'not arbitrarily invented, but given as problems by the very nature of reason itself' (Kant 1998a: 402; CPR A327). From the standpoint of theoretical reason, such a concept may seem empty, since it is 'only an idea' – that is, nothing corresponding to the idea can ever be encountered empirically. But in the practical use of reason, in deciding what to do, such ideas are 'always fruitful in the highest degree' (Kant 1998a: 403; CPR A328), because an idea guides practical reason to 'bring forth what its concept contains', so it 'must serve as a rule, the original and at least limiting condition, for everything practical' (Kant 1998a: 403; CPR A328). As an example, Kant gives 'our idea of perfect humanity' (Kant 1998a: 551; CPR A568), and says, 'Virtue, and with it human wisdom in its complete purity, are ideas' (Kant 1998a: 552; CPR A569). Kant's position is that we have a concept of humanity which is produced by reason rather than experience, and that it includes human rational perfection, and moral goodness or virtue. This idea of humanity does not include empirical features of human beings (their size, physical abilities, colourations, the sound of their voices) but instead the features related to humans as rational beings – knowers and agents. So, to somewhat speculatively fill out Kant's position, the idea of humanity would seem to include the possession of reason in its theoretical use and its practical use. This includes, as aspects of practical reason, the power

of choice and the power to legislate practical moral principles (*Willkür* and *Wille*, to use Kant's later distinction). But beyond this, as an idea of virtue, or properly functioning human reason, it includes the characteristic of placing priority on moral principles over inclination.³ So the idea serves as a practical 'model of virtue' and serves a regulative function as a standard to live up to, and 'it is only by means of this idea that any judgment as to moral worth or its opposite is possible' (Kant 1998a: 396; *CPR* A315).

Kant is surprisingly persistent, throughout many writings of the Critical period, in assigning this regulative, action-guiding role to rationally produced ideas, and specifically to the idea of humanity as human virtue and reason. In a note in *Critique of Practical Reason*, he says that 'moral ideas' ('If I understand by such an idea a perfection to which nothing adequate can be given in experience') serve as 'the indispensable rule of moral conduct and also as the standard of comparison' (Kant 1997a: 106; *C2* 5: 127). Similarly, in *Groundwork*, he explains moral motivation by saying 'here pure reason by means of its Ideas (which furnish absolutely no objects for experience) has to be the cause of an effect admittedly found in experience' (Kant 2002: 259; *G* 4: 460). This repeats his position in *Critique of Pure Reason*, that 'in morality' 'human reason shows true causality', and 'ideas become efficient causes (of actions and their objects)' (Kant 1998a: 397; *CPR* A317). And in the same passage, cited above, in which Kant identifies 'virtue' and 'human wisdom in its complete purity' as ideas, he also says, 'in regard to the principle through which reason places limits on a freedom which is in itself lawless, they can nevertheless serve quite well (if one attends merely to their form) as examples of pure concepts of reason' (Kant 1998a: 552; *CPR* A569). There is, then, a sustained line of thinking in Kant's texts that the power of reason produces an idea of humanity, and that this idea is meant to play an action-guiding role in practical deliberation.

Further examination of *Groundwork* and *The Metaphysics of Morals*, specifically of the passages in which Kant directly discusses humanity, reveal even more direct evidence for taking 'humanity' in the humanity formulation of the categorical imperative to be the rational idea of humanity. Kant's initial presentation of the humanity formulation provides only ambiguous support for this 'rational idea' reading. Most English translations do have Kant saying that, if there is to be a categorical imperative at all, 'it must be such that it forms an objective principle of the will from the idea of something which is necessarily an

end for everyone because it is an end in itself' (Kant 2002: 229; G 4: 428–9), suggesting that what underlies the humanity formulation is an idea of humanity. However, the German word Kant uses for 'idea' here is *die Vorstellung*, not the same word he uses in *Critique of Pure Reason* to describe an idea in the technical sense of a concept provided by reason – that word is *die Idee*. But a few pages later in *Groundwork*, he does closely identify *die Idee* of humanity with the categorical imperative's requirement to treat humanity as an end in itself, equating 'the mere dignity of humanity as rational nature' with 'respect for a mere idea (*eine Idee*)' (Kant 2002: 239; G 4: 439). Similarly, Kant follows this with the claim that the proper object of respect is an 'ideal will, which is possible for us', or, in German, a will that is possible for us 'in der Idee' (Kant 2002: 240; G 4: 440). And in *MM* 6: 451, in explaining why beneficence to oneself is merely a permission rather than a duty, Kant grants that 'lawgiving reason, which includes the whole species (and so myself as well) in its idea of humanity as such, includes me as giving universal law along with all others' (Kant 2002: 200; G 4: 451). These passages, along with others given below as evidence for more specific interpretative points, support taking Kant's position as being that the 'humanity' that plays a central role in the categorical imperative must be an idea of humanity, presented by reason as a concept of perfected human reason and virtue.

The only other apparent candidate that emerges from Kant's Critical writings, for supplying a rationally derived concept of humanity that can play a role in a purely rational moral principle, is what Kant calls an 'ideal' of humanity. This ideal of humanity poses no dire threat to the reading defended above, which takes 'humanity' in the humanity formulation to be the idea of humanity, since the ideal of humanity is itself derived (through reason, not experience) from the idea of humanity, and it also is a concept of humanity as perfected human reason and virtue. The difference is that while an idea is general, a corresponding ideal is given '*in individuo*, i.e., as an individual thing, which is determinable, or even determined, through the idea alone' (Kant 1998a: 551; *CPR* A568). To gain insight into Kant's meaning here, it is useful to turn to *Critique of Judgement*, where Kant explains that '*Idea* properly means a rational concept, and *ideal* the presentation of an individual being as adequate to an idea' (Kant 1987: 80; *C3* 5: 232). Despite being individual and specific, the ideal is still produced by reason rather than being known in experience, but it is a more specific concept than the idea of virtue, inasmuch as it is a concept of what a human being would have to be like in order to

actually be virtuous. Although Kant does sometimes seem to slide between regarding humanity as an idea and as an ideal, the more fundamental rational concept is the idea of humanity, and this seems to be the concept of humanity that is employed in the humanity formulation of the categorical imperative. I will argue below that the corresponding, but more specific, ideal of humanity is better suited to play a role in a rational metaphysics of morals that is based on the categorical imperative.

If, despite the textual evidence, the idea of humanity seems somehow too abstract to be equivalent to the humanity that is an end in itself in the humanity formulation, it may help to consider Kant's discussions of ends in *Groundwork* and *Critique of Judgment*. Kant defines an 'end' as something that serves the will as a ground of its self-determination, or as something that directs one's actions (Kant 2002: 228; G 4: 427). A concept can do this, if it is 'regulative' or action-guiding, and a concept that is presented by reason alone as necessarily action-guiding, would be what Kant calls an 'objective end' or an end in itself. Kant describes the idea of humanity as such a necessarily action-guiding concept, in the passages from *Critique of Pure Reason* in which he defines *Idee*. Kant's discussion of a thing's end (or 'purpose' in some translations, but the same German word, *Zweck*, is translated as either) in *Critique of Judgment* further supports taking the idea of humanity as the end in itself. Kant says that 'insofar as the concept of an object also contains the basis for the object's actuality, the concept is called the thing's purpose', or end (Kant 1987: 20; C3 5: 180). Since the concept of humanity is presented by reason itself as a regulative concept that guides action by demanding that one live up to its standard of perfection, the concept itself is necessarily an end for every rational being. Or, in other words, it is an end in itself.

If the 'humanity' that has a central role in the humanity formulation of the categorical imperative is an idea of perfected humanity that is produced by reason, this has some significant ramifications for our understanding of the principle. For one thing, it means that the starting point for Kant's ethics is not to claim that every actual human individual we encounter is an end in herself. Of course, his ethical system ultimately is meant to provide guidance on how to treat particular humans, but that is an ending point, rather than a starting point, of his overall system, and it requires intervening steps of 'applying' the categorical imperative to experience, steps for which Kant acknowledges the need in developing his 'metaphysics of morals'.

If the humanity formulation is understood as relying on a rational idea of humanity, then recent approaches to understanding the humanity formulation also stand in need of some clarification and revision. Most recent commentators have taken ‘humanity’ in the humanity formulation to refer to ‘a characteristic, or some set of characteristics, of persons’ (Hill 1992: 39). Although there is no unanimity regarding exactly which characteristics (the power to set ends, the capacity for morality, a good will) are the morally relevant ones in marking a person as an end in herself, recent approaches share the assumption that Kantian ‘humanity’ essentially is some set of characteristics possessed by humans.⁴ But taking humanity as a rational idea adds a layer of depth, and an additional requirement, to this claim. What is added by taking humanity to be a rational idea is the explanation of which characteristics are the relevant ones, and why. The relevant characteristics, if we take humanity to be an idea of reason, presumably are the power to set ends, the possession of theoretical reason and understanding, the capacity to legislate moral principles to oneself and the prioritizing of moral principles over inclination. The reason these are the relevant characteristics is not because we observe empirically that some, all or most humans have these traits, nor because our moral intuitions tell us they are what makes a person morally significant. It is because the only concept of humanity that can feature centrally in a purely rational moral principle is the rationally produced idea of humanity, which is an idea of perfected humanity and virtue.

Of course, defenders of other readings of ‘humanity’ may resist taking humanity to be a concept of perfected human reason and virtue, but their resistance does seem to require rejecting the entire approach of taking ‘humanity’ in the humanity formulation as being equivalent to the rationally produced idea of humanity. It is not only that Kant consistently says, in *Critique of Pure Reason*, that this idea of humanity is a concept of human virtue, nor is it just that no other rationally produced concept of humanity is mentioned, or obviously available, in Kant’s philosophy (except for the rational ‘ideal’ of humanity which is derived from the idea). There also is a Kantian rationale for thinking that a concept of humanity delivered by reason actually must be a concept of perfected, properly functioning human reason. Kant maintains that experience can never tell us what belongs necessarily to a concept (Kant 1996: 18; *MM* 6: 226–7). He says this in order to explain why freedom cannot be defined as freedom to choose between moral and immoral action, and adds: ‘Only freedom in relation to the internal lawgiving of reason is really an ability; the possibility of deviating from

it is an inability.’ Reason can only deliver to us a concept of a properly functioning rational being who acts in the ways reason demands. It is true that ‘experience proves often enough’ that people also can act contrary to what is rationally (and so, morally) required, but ‘we still cannot comprehend how this is possible’ (Kant 1996: 18; *MM* 6: 226).⁵ Reason can only tell us what a proper or perfected human reason is, and the fact that particular human beings fall short of this is only known by experience.

Elsewhere in *The Metaphysics of Morals*, Kant appears to remind the reader that the concept of humanity underlying all duties is a concept of perfected human virtue. To explain why an individual must leave punishment of wrongdoers to the legal system, rather than undertaking it privately, Kant says that ‘in ethics’ we must ‘regard human beings as in a rightful condition but in accordance only with laws of reason (not civil laws)’ (Kant 1996: 207; *MM* 6: 460). On the next page, he expands on this idea, saying that, although we can observe individual humans’ moral shortcomings, this ‘does not justify attributing to them a predisposition to these vices belonging to their species, any more than the stunting of some trees in a forest is a reason for making them a special kind of plant’ (Kant 1996: 208; *MM* 6: 461). His concern here is not an accurate empirical description of humanity as a biological species, but on obtaining an accurate rational concept of humanity. He says ‘any vice, which would make human nature itself detestable, is inhuman when regarded objectively. But considered subjectively, that is, in terms of what experience teaches us about our species, such vices are still human.’ And he explains that we cannot regard humans as both virtuous and vicious, since ‘dividing something into two heterogeneous things yields no definite concept at all’ (Kant 1996: 208; *MM* 6: 461). Passages like these, along with the *Groundwork* claim already cited above that the proper object of respect is an ‘ideal will, which is possible for us’ (Kant 2002: 240; *G* 4: 440), are evidence of Kant’s sustained commitment to the rational idea of perfected human reason and virtue as being the concept of humanity in the humanity formulation, and so as the concept of humanity at the heart of his ethical system.

2. Humanity as an Ideal in the Purely Rational (?) Metaphysics of Morals

Even if the categorical imperative is meant to be a purely rational principle, there is still room to wonder whether Kant means the entire metaphysics of morals, which is comprised of not only the categorical

imperative but also the more detailed system of duties based on that underlying principle, to be exclusively a product of reason. Although it sometimes appears that Kant is applying the (rationally produced) categorical imperative to empirical circumstances in *The Metaphysics of Morals*, I will argue that we should take seriously his own claims that the entire metaphysics of morals is purely rational, and does not take empirically known features of humans into account. If this is right, then only rational concepts of humanity must figure in the metaphysics of morals. I propose that the only concepts of humanity that play a role in the metaphysics of morals are the rational idea of perfected human virtue, and the derivative, more specific, rational concept that Kant calls an ‘ideal’ of humanity.

There is certainly *prima facie* textual evidence that Kant means the metaphysics of morals to be purely rational, and so means it, like the categorical imperative on which it is based, to employ only rationally produced concepts of humanity. Kant says that the ‘empirical part’ of ethics, or its application to human individuals, should be given the title ‘practical anthropology’ (Kant 2002: 190; G 4: 388) and he goes on to reiterate that it is ‘a matter of utmost importance, to forge for once a pure moral philosophy, completely cleansed of everything that may be only empirical and that really belongs to anthropology’ (Kant 2002: 191; G 4: 389). So, ‘Pure philosophy (metaphysics) must therefore come first, and without it there can be no moral philosophy at all’ (Kant 2002: 192; G 4: 390). In chapter 2 of *Groundwork*, he reiterates that a metaphysics of morals is meant to be ‘found completely a priori and free from empirical elements in concepts of pure reason and absolutely nowhere else to the slightest extent’ and that such a ‘completely isolated metaphysics of morals, mixed with no anthropology, no theology, no physics or hyperphysics’, is an ‘indispensable underlying support for all theoretical and precisely defined knowledge of duties’ (Kant 2002: 211–12; G 4: 410). He repeats two pages later that in ethics, ‘pure philosophy’ or ‘metaphysics’ must be ‘expounded independently’ of all anthropology (Kant 2002: 213; G 4: 412). Kant presents a consistent overall picture in *Groundwork*, that moral philosophy has two parts, a purely rational part that he calls a ‘metaphysics of morals’, and an empirical application of the rules of the metaphysics of morals to actual human beings, which he calls ‘anthropology’.

But it may appear that his views, or at least his terminology, change between *Groundwork* and *The Metaphysics of Morals*. In the latter work, Kant often includes details of human life, including elements like

the existence of shops and commerce, the effects of consuming alcohol, and the practice of selling one's hair and teeth. He also devotes considerable space to the treatment of criminals and wrongdoers, which is hardly compatible with viewing humans only as perfectly rational and virtuous (e.g. Kant 1996: 104–10; *MM* 6: 331–7, 210, 463). Passages like these naturally make it tempting to think empirical observation plays a role in Kant's metaphysics of morals, and to read his own descriptions of his project as allowing that possibility. For example, near the beginning of the book, he says that

a metaphysics of morals cannot dispense with principles of application, and we shall often have to take as our object the particular nature of human beings, which is cognized only through experience, in order to show in it what can be inferred from universal moral principles. (Kant 1996: 10; *MM* 6: 216–17)

One plausible reading of this passage is that the 'principles of application' which take empirically observed features of humans into account are actually part of the metaphysics of morals, which implies that Kant has abandoned the claim that a metaphysics of morals is purely rational. And when Kant says, later in the book, that 'practical philosophy' is like other philosophy in needing to be based on 'a system of pure rational concepts' or 'metaphysical first principles', it may be thought that Kant is only identifying the various formulations of the categorical imperative as purely rational 'first principles', with the rest of the metaphysics of morals incorporating empirical observation (Kant 1996: 141; *MM* 6: 375).

But this impression, that Kant's metaphysics of morals is an application of rational principles to empirically known human circumstances, is mistaken. The particular passages that seem to support this reading actually are consistent with taking a metaphysics of morals to be purely rational, and, more importantly, the overall evidence from within *The Metaphysics of Morals* supports the 'purely rational' reading.

The passages that seem at first glance to support taking the metaphysics of morals as an amalgam of rational principles and empirical application prove, on closer examination, to provide more support for the 'purely rational' reading. In *The Metaphysics of Morals*, 6: 217, when Kant says that a metaphysics of morals also needs empirically influenced principles of application, these principles of application might be

taken either as a part of the metaphysics of morals or as a necessary accompaniment to it. But the rest of the passage strongly supports the latter reading. Kant begins the paragraph by defining a metaphysics of morals as ‘a system of a priori cognition from concepts’, and ends the paragraph by saying that ‘a metaphysics of morals can not be based upon anthropology but can still be applied to it’ (Kant 1996: 10; *MM* 6: 216–17). Here, Kant is adhering to his distinction between a rational metaphysics of morals and an empirical ‘anthropology’ or application to actual human beings, and the principles of application are best taken to be part of anthropology. Similarly, in *The Metaphysics of Morals*, 6: 375, when Kant asks whether practical philosophy needs purely rational first principles, his answer is that it does, and these first principles are not just the various formulations of the categorical imperative, but the more complete system of duties catalogued in *The Metaphysics of Morals*. In the opening lines of the passage, Kant identifies these ‘metaphysical first principles’ as comprising a ‘metaphysics’ or ‘a system of pure rational concepts’ (Kant 1996: 141; *MM* 6: 375). Even more decisively, it should be remembered that the titles of the two halves of the book are *Metaphysical First Principles of the Doctrine of Right* and *Metaphysical First Principles of the Doctrine of Virtue*, indicating that Kant is calling the contents of a metaphysics of morals ‘metaphysical first principles’, which he also says are rational rather than empirical (Kant 1996: 210–11; *MM* 6: 375–7).

Taking these passages as support for thinking that a metaphysics of morals is purely rational renders them consistent with the bulk of textual evidence in *The Metaphysics of Morals*. Kant says that he is excluding empirically known differences among humans because his main concern in the book is describing ‘metaphysical first principles of a doctrine of virtue’ and that this project ‘has to do only with its pure rational principles’ (Kant 1996: 213; *MM* 6: 468). And, in a passage that virtually forces us to view the ‘principles of application’ mentioned in *MM* 6: 217 as a complement to a metaphysics of morals rather than as a part of it, near the end of the book Kant says

just as a passage from the metaphysics of nature to physics is needed – a transition having its own special rules – something similar is rightly required from the metaphysics of morals: a transition which, by applying the pure principles of duty to cases of experience, would schematize these principles, as it were, and present them as ready for morally practical use. (Kant 1996: 214; *MM* 6: 468)

Kant then specifies that these principles of application ‘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 1996: 214; *MM* 6: 469). This fits also with an earlier statement that ‘every morally practical relation to human beings is a relation among them represented by pure reason’ (Kant 1996: 200; *MM* 6: 451). The overall textual evidence strongly suggests that Kant means the entire metaphysics of morals, not just the categorical imperative on which it is based, to be purely rational.

One remaining obstacle to accepting this reading of the metaphysics of morals as purely rational is Kant’s habit of including empirical details of human life in *The Metaphysics of Morals*, including knowledge of the moral failings of human beings. But there is an obvious explanation of this habit, which is that Kant is not assiduously limiting the contents of *The Metaphysics of Morals* to just what is a part of the metaphysics of morals proper, but also is including additional comments for purposes of clearer or more compelling presentation. This is consistent with the way he presents many of the specific moral issues that arise from experience (such as the moral status of heroic martyrdom, alcohol consumption or white lies), labelling them as ‘Casuistical Questions’, and generally leaving these questions unanswered, seemingly as issues to be resolved in later empirical application or ‘anthropology’. Empirical details also come up in other parts of the book, not only in Kant’s casuistical questions, but they also may be regarded as details that are mentioned in *The Metaphysics of Morals*, but are not strictly part of a metaphysics of morals. To take Kant to be including empirical details in the book does not even require thinking of him as guilty of any slip – if some empirical illustrations and questions help to clarify the purely rational system of duties, there is no obvious reason he cannot consistently include them in the book, even if elucidating the purely rational system of duties is the main point.

For this reason, Kant’s inclusion of the treatment of wrongdoers as a topic does not show that the metaphysics of morals is not meant to be rationally produced, even if the concept of humanity that is provided by reason must include human virtue, not vice. When Kant says that a vicious person must be treated with respect despite failing to deserve it, he may well be adding the case of the vicious person as a contrast, to emphasize that the (virtuous) human as rationally conceived in a metaphysics of morals must be thought of as deserving respect – ‘respect for a human being as a moral being (holding his duty in highest esteem)’

(Kant 1996: 210; *MM* 6: 464). And at least at the level of explaining basic principles, Kant takes pains to say that our awareness of the need for laws and punishment does not arise from empirical knowledge of wrongdoing. 'It is not experience from which we learn of human beings' maxim of violence and of their malevolent tendency to attack each other ... It is therefore not some fact that makes coercion through public law necessary' (Kant 1996: 89; *MM* 6: 312). Instead, lack of security and the possibility of violence lie 'a priori in the rational idea' of the right to act on private, subjective principles in a state of nature. So, Kant seems to emphasize that a metaphysics of morals is supposed to be purely rational, by making it a point to explain that the need for a doctrine of right does not arise from empirical observation, and his less central discussions of how to treat wrongdoers may reasonably be regarded as lying outside the scope of a metaphysics of morals proper, despite being included in the book of that name.

If Kant does mean the entire metaphysics of morals to be strictly rational, then the concept or concepts of humanity that he includes when developing the system of duties must be purely rational as well. The rationally produced idea of humanity as perfect human reason and virtue plays a central role, since it is the concept of humanity that Kant employs in the humanity formulation of the categorical imperative, and the whole system is based on the categorical imperative. And in many passages in *The Metaphysics of Morals*, cited above, Kant does talk about an idea of humanity presented by reason. But Kant gives some signs that in moving from the basic principle to more specific duties, he also shifts focus from the idea of humanity to a correlative 'ideal' of humanity.

An ideal, like an idea, is a concept produced by pure reason (Kant 1998a: 552; *CPR* A569–70, 554, 573–4, 556, 576). In fact, an ideal is a more specific concept derived through pure reason from a corresponding idea. Kant says an ideal is 'the concept of an individual object that is thoroughly determined merely through the idea' (Kant 1998a: 554; *CPR* A574). An ideal is 'a concept thoroughly determined a priori' from an idea, in that the idea tells us what other characteristics must be predicated of an object in order for it to be adequate to the idea – 'an idea, as an original concept, excludes a multiplicity of predicates, which, as derived through others, are already given, or cannot coexist with one another' (Kant 1998a: 554; *CPR* A573–4). As an example of this, Kant says that our ideal of humanity contains more than our idea of 'perfect humanity', since the idea of perfect humanity includes only

‘all those properties belonging essentially to this nature’ of human perfection, but the ideal also contains ‘everything besides this concept that belongs to the thoroughgoing determination of this idea’ (Kant 1998a: 551; *CPR* A568). That is, the ideal of perfected human wisdom and virtue tells us what an individual human would have to be like in order to be rationally perfect. She would have to ‘possess everything that our reason links with the morally good: goodness of soul, or purity, or fortitude, or serenity, etc.’ (Kant 1987: 84; *C3* 5: 235). This ideal of a good human being, like the corresponding idea of human virtue, is never to be met with in experience, but both have ‘practical power (as regulative principles)’ (Kant 1998a: 552; *CPR* A569). Kant says ideals ‘supply reason with a standard which is indispensable to it, providing it, as they do, with a concept of that which is entirely complete of its kind’ (Kant 1998a: 552; *CPR* A569–70).

Because the ‘ideal’ of humanity in Kant’s technical sense supplies us with a standard for virtue, it also accords with the more everyday sense of an ideal as a model for action. When introducing the ideal of humanity in *Critique of Pure Reason*, Kant says that this ideal is a concept of an individual human that serves as an ‘archetype’ of virtue and that ‘we have in us no other standard for our actions than the conduct of this divine human being, with which we compare ourselves, judging ourselves and thereby improving ourselves, even though we can never reach the standard’ (Kant 1998a: 552; *CPR* A569). Kant continues this line of thought in *Religion Within the Limits of Reason Alone*, saying, ‘Now it is our universal human duty to elevate ourselves to this ideal of moral perfection, i.e. to the prototype of moral disposition in its entire purity’ (Kant 1998b: 80; *R* 6: 61). Kant does not hesitate to identify this ideal as ‘an ideal of humankind’ (Kant 1998b: 84; *R* 6: 66) and ‘an ideal of humanity pleasing to God (hence of such moral perfection as is possible to a being pertaining to this world and dependent on needs and inclinations)’ (Kant 1998b: 80; *R* 6: 61). Even when Kant acknowledges Jesus as a ‘prototype’ of moral perfection, he adds that this prototype really ‘resides in our own morally-legislative reason’ and ‘is present as model already in our reason’ (Kant 1998b: 81; *R* 6: 62). This echoes Kant’s statements in *Critique of Pure Reason* that the wise man of the Stoics is ‘an ideal, i.e., a human being who exists merely in thought’ (Kant 1998a: 552; *CPR* A569).

An ideal of humanity, a concept of what a human being would have to be like in order to be rational and virtuous, which is derived from the more general idea of perfect humanity, is a concept of humanity that

Kant deploys along with the idea of humanity in developing a metaphysics of morals. Some of Kant's own descriptions of his project in developing a metaphysics of morals suggest that he begins with an idea of humanity in the categorical imperative, but then supplements it with a more specific ideal of humanity that includes the feelings and traits that a human would need in order to match the idea of perfected human rationality and virtue.

Most notably, Kant apparently is filling out the ideal of humanity when, near the beginning of *The Metaphysical First Principles of the Doctrine of Virtue*, he describes 'certain moral endowments' that 'lie at the basis of morality, as subjective conditions of receptiveness to the concept of duty' (Kant 1996: 159; *MM* 6: 399). Kant says that 'every human being has them, and it is by virtue of them that he can be put under moral obligation'. The four necessary conditions for human moral action and virtue that Kant describes here are 'moral feeling' (the 'susceptibility to feel pleasure and displeasure merely from being aware that our actions are consistent with or contrary to the law of duty'), 'conscience', 'love of human beings' (as an 'inclination to beneficence') and 'respect' (*Achtung*, or respect for moral law and so for oneself as a moral legislator) (Kant 1996: 160–2; *MM* 6: 399–403). These characteristics belong to the rational concept of humanity as an ideal, since we know through reason alone that a human being would have to possess them in order to live up to the idea of human virtue. Kant says: 'Consciousness of them is not of empirical origin; it can, instead, only follow from consciousness of a moral law, as the effect this has on the mind' (Kant 1996: 159; *MM* 6: 399).

Kant's statements about what is excluded from the concept of humanity in a metaphysics of morals also fit with taking humanity as a rational, rather than empirically acquired, concept. Near the very end of *The Metaphysics of Morals*, Kant says that the entire work he has just completed, the metaphysics of morals, has not taken into account any empirically observed differences between humans, including differences of age, gender, social status, education and moral development, because such differences do not 'proceed a priori from rational concepts', and he says that taking such differences into account is just a matter of adding 'corollaries' to a metaphysics of morals (Kant 1996: 214; *MM* 6: 468–9). If even individuals' states of 'moral purity and depravity' are to be left out of consideration, this is not because moral virtue is irrelevant to the concept of humanity being employed, but because the only rational concepts of humanity that are available are concepts of

perfected human reason and virtue, and because the role of the ideal of humanity is not diminished by empirically observed failures of individual humans to live up to the ideal. Kant tells us this explicitly, saying ‘Any high praise for the ideal of humanity in its moral perfection can lose nothing in practical reality from examples to the contrary, drawn from what now are, have become, or will presumably become in the future’ (Kant 1996: 164; *MM* 6: 406). He then immediately reiterates his distinction between a rational system of morals and empirically based application, saying, ‘anthropology, which issues from mere empirical cognition, can do no damage to anthroponomy, which is laid down by reason giving laws unconditionally’. This echoes Kant’s statement near the beginning of the book that a metaphysics of morals is distinct from ‘moral anthropology’ and that the empirically known nature of human beings ‘will in no way detract from the purity of these principles [of a metaphysics of morals]’ (Kant 1996: 10; *MM* 6: 217). The exclusion of empirically observed individual human differences from the metaphysics of morals also is consistent with *MM* 451, where he says that ‘lawgiving reason’ is concerned with the ‘idea of humanity as such’, and ‘not the human being’ (*ihrer Idee der Menschheit überhaupt... nicht der Mensch*) (Kant 1996: 200; *MM* 6: 451).⁶ There is strong textual evidence that Kant really does mean the metaphysics of morals to be employing only concepts of humanity delivered by reason alone, namely the idea and corresponding ideal of humanity as human virtue.

Kant’s development of specific categories of duty in the metaphysics of morals further reinforces this interpretation of the metaphysics of morals as employing the rational idea and ideal of humanity.

Kant’s account of the duty of self-perfection provides the most obvious support for taking humanity as a rationally produced concept of perfected humanity. In *The Metaphysics of Morals*, Kant describes self-perfection as one of the basic categories of duty (Kant 1996: 150–1; *MM* 6: 386–7), and follows this with a more detailed further discussion of the duties of natural and moral self-perfection (Kant 1996: 194–7; *MM* 6: 444–7). These discussions expand on his position in *Groundwork* that we have a duty to develop our talents (Kant 2002: 223–4; *G* 4: 422–3). It is first worth noting that the justification of a duty of self-perfection does seem to require some concept of perfected humanity. If humanity is instead taken to be just some set of characteristics possessed by typical rational humans, as in standard readings of the humanity formulation, then it is hard to see how a duty of self-perfection follows. If humanity is just the power to set ends, or if it is that plus the possession of theoretical reason, or if it is a

mere potential (not necessarily a realized potential) for moral action, then the humanity formulation is attributing an incomparable dignity and special status as an end in itself to some trait that all minimally rational adult humans already possess. Then there is no apparent rational room to demand change. The characteristics and potentials (even if unrealized) that one already has are enough to confer incomparable worth. In contrast, if the humanity that is an end in itself is some perfected version of humanity, then it can serve as an ideal towards which one ought to strive. And this is, of course, how Kant describes the practical function of the ideal of humanity as a rational ideal or ‘archetype’. This fits with Kant’s statement that someone’s duty of self-perfection is ‘a duty to make his end the perfection belonging to a human being as such (properly speaking, to humanity)’ (Kant 1996: 150; *MM* 6: 386).

In fact, a closer examination of this passage suggests that Kant is relying on his own technical definition of an ‘ideal’ from *Critique of Pure Reason*, in order to develop the duty of self-perfection. Kant clarifies a possible ‘misinterpretation’ of the perfection that is required (Kant 1996: 150; *MM* 6: 386). He denies that the concept of perfection involved in this duty is ‘a concept belonging to transcendental philosophy’, which is a ‘quantitative perfection’ which ‘can be only one (for the totality of what belongs to a thing is one)’. Instead, the perfection that is ‘under discussion’ in the duty of self-perfection is ‘a concept belonging to teleology’, of ‘qualitative (formal) perfection’, and specifically a concept of ‘the harmony of a thing’s properties with an end’, and ‘one thing can have several qualitative perfections’.⁷ In this passage, Kant is denying that the mere completeness or ‘totality of the manifold which, taken together, constitutes a thing’ is an adequate concept of perfection to play a practical role in justifying the duty of self-perfection. Kant presumably has in mind, partly, a denial that the type of ontological perfectionism proposed by Leibniz and Wolff is a sound basis for developing moral duties. For both Leibniz and Wolff, perfection is a matter of the greatest number of parts combined in the simplest possible ways, which constitutes the greatest possible degree of reality (Leibniz 1969; Wolff 1976, 1983).⁸ But Kant also seems to be ruling out his own transcendental idea of humanity as an adequate basis for the duty of perfection, since this rational idea is distinguished by its ‘completeness’ and ‘systematic unity’ (Kant 1998a: 551; *CPR* A567–8). But the ideal of humanity specifies a number of particular perfections that would have to be possessed in order for a human being to be adequate to the idea of perfect humanity (perfected human reason and virtue).

So, even if the ideal of humanity is not fully achievable by any actual human, it is well suited to serve as a model or archetype for human perfection, or for ‘cultivating one’s faculties’ of ‘understanding’ and of ‘will’ (Kant 1996: 150–1; *MM* 6: 387).

Kant’s discussion of duties of respect for others also accords well with taking humanity as an ideal of reason, despite a possible first impression to the contrary. The contrary first impression suggests that, for Kant, duties of respect are founded initially on recognizing an incomparable worth or dignity in every individual human person, as encountered in experience, and then treating each person with respect because of that dignity. The passages suggesting this picture include Kant’s claim that everyone must ‘acknowledge, in a practical way, the dignity of humanity in every other human being’, because ‘Every human being has a legitimate claim to respect from his fellow human beings and is in turn bound to respect every other’ (Kant 1996: 209; *MM* 6: 462), and that each person ‘possesses a dignity (an absolute inner worth) by which he exacts respect for himself from all other rational beings in the world’ (Kant 1996: 186; *MM* 6: 435). These passages, along with *Groundwork* 4: 428, where Kant first introduces the idea of humanity as an end in itself and says that ‘Rational beings, on the other hand, are called persons because, their rational nature already marks them out as ends in themselves ... and consequently imposes restrictions on all choice making (and is an object of respect)’ (Kant 2002: 229; *G* 4: 428), may seem to support a familiar picture of Kant’s ethics, which starts with every individual’s fundamental respect-worthiness. But this picture is undermined by some of Kant’s other statements in *The Metaphysics of Morals* about respect. In *MM* 6: 463, Kant reaffirms that we ought to treat every person with respect, but says that this is despite the fact that some individuals are unworthy of respect. He says, ‘I cannot deny all respect to the vicious man as a human being; I cannot withdraw at least the respect that belongs to him in his quality as a human being, even though by his deeds he makes himself unworthy of it’ (Kant 1996: *MM* 6: 463). Reinforcing this idea that not everyone deserves respect, three pages later Kant says that ‘examples of respect that we give others can arouse their striving to deserve it’, which again implies that some people are not yet worthy of respect (Kant 1996: 212; *MM* 6: 466).⁹ Apparently, the respect-worthiness of every individual human being is not the basis of the duty of respect, since not every human is worthy of respect.

Instead, the metaphysics of morals, including the duty of respect, takes as its object a rationally given concept of humanity. This humanity is a

perfected human nature, which is deserving of respect, and the duty of respect is of ‘respect for a human being as a moral being (holding his duty in highest esteem)’ (Kant 1996: 210; *MM* 6: 464). Empirical observations of individuals who fall short of the ideal or are in a state of ‘depravity’ (Kant 1996: 214; *MM* 6: 468) are a matter for later casuistry or ‘anthropology’. So the otherwise paradoxical claim that we ought to treat every human with respect despite the fact not everyone is worthy of respect can be explained by keeping in mind that what we really must respect is humankind as it falls under the rational concept of humanity. This fits well with Kant’s repeated claim that his metaphysics of morals describes duties to ‘humanity as such’, rather than to specific types of human individuals. Kant actually uses two different phrases which are both translated as ‘humanity as such’. Sometimes he speaks of human beings or humanity *als solcher*, which is most straightforwardly translated as the English ‘as such’, with the emphasis being that it is humanity itself, or humanity in virtue of its own characteristics, which must be treated as an end in itself (Kant 1996: 210, 214; *MM* 6: 464, 468). But he also, more frequently, says it is humanity *überhaupt* which is the basis of moral duties, and *überhaupt* has more of a connotation of humanity ‘generally’ or ‘overall’ rather than instantiated in specific individuals (Kant 1996: 150, 157, 200, 212, 213; *MM* 6: 386, 395, 451, 466, 468). By itself, the phrases ‘humanity as such’ or ‘humanity in general’ leave unanswered the question of where the general concept of humanity is supposed to come from, if not from experience, But Kant’s technical accounts of the rational concepts of humanity as an idea and an ideal supply the answer.

If Kant really does mean the metaphysics of morals to be derived through reason alone, and to employ only rationally produced concepts of humanity, this renders him consistent in maintaining a distinction between the metaphysics of morals and ‘moral anthropology’ or the application of that system of principles to real, individual humans. But he simply never gets to a thorough discussion of moral anthropology. This suggests that Kant underestimates the complexities and potential controversies in moving from ideal to non-ideal theory (to adapt Rawls’s terms), but Kant at least does give indications of what his moral anthropology would attempt to do. After emphasizing the division of metaphysics from anthropology, he says that moral anthropology

would deal only with the subjective conditions in human nature that hinder people or help them in fulfilling the laws of a metaphysics of morals. It would deal with the development,

spreading, and strengthening of moral principles (in education in schools and in popular instruction), and with other similar teachings and precepts based on experience. (Kant 1996: 10–11; *MM* 6: 217)

The point of applying the metaphysics of morals, then, seems not to be to alter the basic moral requirements it contains, to fit with individual human differences, but rather to take those moral requirements as fixed and to attempt to promote adherence to them. By encouraging the acceptance of these moral requirements, anthropology also would aim at the moral improvement of humankind, or the approximation of humans to the ideal of humanity. Although one cannot ensure that another person will act rightly or have a good character, one can seek to promote the general conditions that will tend to lead humans to moral improvement. This is consistent with Kant's claims about the overall purpose of education, in *Lectures on Pedagogy*, where he maintains that education should aim at fulfilling humanity's destiny and cultivating a human predisposition to perfection (Kant 2007: 437; *LP* 9: 441, 438, 442, 439–43, 444–9). In these lectures, he often refers to an idea and ideal of humanity, such as when he says education should be conducted 'in a manner appropriate to the idea of humanity and its complete vocation' (Kant 2007: 442; *LP* 9: 447), or that each person 'has an ideal of humanity before his eyes' and that he reprimands himself when he 'compares himself' with this idea and falls short of it (Kant 2007: 476; *LP* 9: 489).

3. Further Textual Considerations

Taking 'humanity' in the humanity formulation of the categorical imperative as an idea of human perfection, and in the metaphysics of morals as this idea plus the accompanying rational ideal, no doubt has both interpretative advantages and disadvantages. Here, I will only quickly describe two interpretative advantages.

Taking the humanity that is an end in itself to be a rationally produced concept also makes sense of Kant's frequent way of describing the kingdom of ends as an idea or an ideal. The 'ends' in the kingdom of ends are meant to be ends in themselves (Kant says the kingdom of ends is a 'systematic union of rational beings through shared objective laws', all of them 'under the law that each of them should treat himself and all the others never merely as means but always at the same time as an end in itself': Kant 2002: 234; *G* 4: 433), so if both the end in itself and the

kingdom of ends are concepts of pure reason, this helps to render Kant's texts consistent. In *Groundwork*, 4: 439, Kant calls the kingdom of ends an idea, and later he not only speaks of the idea of a purely intelligible world (as opposed to the natural world), but also says this idea gives rise to 'a splendid ideal of a universal kingdom of ends' (Kant 2002: 261; G 4: 462). This echoes what he says when he introduces the concept of a kingdom of ends, that a kingdom of ends is 'admittedly only an ideal' (Kant 2002: 234; G 4: 433). There is strong textual evidence, then, that Kant takes the kingdom of ends, and so also the ends in themselves that comprise this ideal realm, to be concepts delivered by pure reason.

Taking humanity as an idea also reinforces and deepens our comprehension of Kant's talk of two worlds, one sensible and one intelligible, and of the corresponding distinction between humans as *homo phenomenon* and *homo noumenon*. Building on his distinction from *Critique of Pure Reason* between things-in-themselves and empirical objects, Kant spends much of chapter 3 of *Groundwork* contrasting an 'intelligible world' or 'world of understanding' with a 'sensible world' or world of 'appearances' (Kant 2002: 250–62; G 4: 451–63). Kant's position is that, when viewed as a member of the intelligible world, a human being must take herself to be free of the laws of empirical causation which govern the sensible world, and the only way to view herself as free in this way is to think of herself as providing herself with moral laws, which are reasons for action independent of empirical influences. But each of us must also view herself and other humans as part of the natural or 'sensible' world, which means we are subject to natural, causal laws and so are prone to act on inclinations. Viewed as part of the intelligible world, a human is what Kant calls *homo noumenon*, or a being of pure reason, while viewed as part of the ordinary world of the senses, a human is *homo phenomenon*, a physical part of the physical world. The moral principles which one legislates to oneself as reasons for action as *homo noumenon* apply regardless of the causal laws and restrictions of the sensible world, which leads Kant to say that a human viewed as a being 'endowed with inner freedom (*homo noumenon*)' can be thought of as putting herself under obligation, because as *homo phenomenon*, she must obey the laws she legislates as *homo noumenon* (Kant 1996: 174; MM 6: 418). The rational idea of humanity, which includes perfect human virtue, is in fact the same concept as *homo noumenon*. And it is *homo noumenon* that is an end in itself. While a human as *homo phenomenon* 'shares with the rest of the animals, as offspring of the earth, an ordinary value'

a human being regarded ‘as a person, that is, as a subject of a morally practical reason, is exalted above any price; for as a person (*homo noumenon*) he is not to be valued merely as a means to the ends of others or even to his own ends, but as an end in himself’ (Kant 1996: 186; *MM* 6: 434–5). Although some sceptics will still find the picture spooky, taking humanity as an idea provides a further, deflationary explanation of what it means to view humankind from two standpoints, as both a rational being and a natural being. We can think of humanity either as an idea of perfected humanity provided by reason, or as we encounter it in experience, in the form of human individuals, but the rational idea has priority inasmuch as it provides a regulative idea or standard to which we as actual individuals must comply.¹⁰

4. Postscript: Response to Denis

The view defended in this paper is, I think, largely compatible with a position that I have defended elsewhere, that ‘humanity’ in the humanity formulation of the categorical imperative is best read as ‘good will’, and so it is good will that must be treated as an end in itself.¹¹ Taking humanity as a rationally produced idea of perfected wisdom and virtue reinforces that earlier position, and also suggests a response to an objection that Lara Denis has offered to the position (Denis 2010).

I do not pretend to capture all the details of Denis’s article here, but I will offer a brief response to the objection that Denis calls her ‘main argument against Dean’s position’ (Denis 2010: 120). Her argument relies on taking humanity to be some ‘aspect’ of the self that places a human agent under moral obligations, and Denis cites several passages in which Kant identifies the aspect of a person which imposes duties or obligations on a person (the ‘obligans’) as ‘humanity’ (Denis 2010: 128–30, 132–4). She also notes that ‘all human agents are subjects of obligation’ (each is an ‘obligatus’) (Denis 2010: 127). From the premises that all human agents are subjects of obligation and that the source of this obligation must be the agent’s own humanity, Denis concludes that humanity must be ‘an aspect of (something “in”) all human agents’ (Denis 2010: 128). And since a good will is not an aspect of or something ‘in’ all human agents, it follows that ‘a good will ... is not humanity’, so my reading of the humanity formulation is mistaken (Denis 2010: 128).

I grant that Kant sometimes says that it is one’s own humanity which imposes duties or moral obligations upon one, and grant that this is one way to capture his overall settled view about the source of obligations.

I also grant that he thinks all human agents are subject to moral obligation. Nevertheless, this does not support Denis's sub-conclusion that humanity must be something 'in' all human agents. That inference relies on mistakenly taking Kantian humanity (the same humanity which is an end in itself) to be some characteristic or property that is possessed (or lacked) by actual, individual humans. Although this view of humanity as a characteristic has been common in commentaries on Kant's ethics, Kant's account of how a human being places herself under moral obligations strongly suggests the alternative reading I have defended above, in section 1, namely that humanity instead is a rational concept that plays a necessary role in moral thinking. It is not that one part of a person, her humanity, places another part of her under moral obligation. Instead, it is that a person thought of in one way, as *homo noumenon*, places herself as thought of in another way, as *homo phenomenon*, under moral obligation. To view a human being as the source of moral obligations is also to view her as *homo noumenon*, and so to view her as an intelligible being who acts in the ways that rationality demands – in other words, it is to view her as falling under the rational concept of humanity as an idea of perfected human reason and virtue. The humanity that is an end in itself is humanity as conceived in this way, and when thinking of duties, either duties to humanity or duties as imposed by an agent's own humanity, this is how we must think of humanity.

Even some of the passages that Denis emphasizes in describing the 'humanity' that is the source of obligations actually support this reading of humanity as a rational concept of perfected human reason and virtue. Of the passages she cites in support of taking humanity to be the source of obligations, she devotes the most space to Vigilantius's notes on Kant's lectures on ethics, in which Vigilantius actually records Kant as saying that we must view the human being both as an agent and as a law-giver, and that as a law-giver, we must think of him 'first of all, as an ideal, as he ought to be and can be, and call this idea *homo noumenon*' and that this ideal is 'only a personified idea' which places the human as *homo phenomenon* under obligation (Kant 1997b: 341; *LE* 27: 593, cited in Denis 2010: 129 and 132). She also says that 'the notion of a pure, non-empirical will ... is at the core of Kant's notion of humanity as inner *obligans*' (Denis 2010: 134), and cites as evidence Kant's discussion of a pure will, or 'one that would be completely determined from a priori principles alone, without any empirical motives' (Kant 2002: 192; *G* 4: 390). She describes these as 'passages in which Kant portrays humanity qua inner *obligans* as an archetype of moral perfection' (Denis 2010: 134). If so, this supports taking the

concept of humanity that is central in Kant's ethics to be a concept of perfected human reason and virtue, rather than some more minimal feature of rationality.¹²

Notes

- 1 Kant 2002: 190; G 4: 388, 191, 389, 208–13, 406–11; Kant 1996: 141; MM 6: 375, 214, 468; and several other passages cited below. I will usually cite Kant's work parenthetically, using the following abbreviations: CPR = *Critique of Pure Reason*, C2 = *Critique of Practical Reason*, C3 = *Critique of Judgment*, G = *Groundwork for the Metaphysics of Morals*, LP = *Lectures on Pedagogy*, LE = *Lectures on Ethics*, MM = *The Metaphysics of Morals*, R = *Religion Within the Limits of Reason Alone*.
- 2 For some of the textual references here, and for the connection between Kant's technical use of 'idea' and humanity as an 'ideal' in the less technical sense, I am indebted to Richard Smyth's very short paper, 'Why Respect Humanity?' which to the best of my knowledge remains unpublished.
- 3 See Kant 1998b: 58–9, R 6:36–7, for more detail on Kant's account of good and evil moral character.
- 4 Some have emphasized the power to set ends or the power to organize ends into a coherent whole as the most morally essential feature of humanity (Korsgaard 1996: 17, 106–32, 346; Wood 1999: 118–20), others have included a wider set of features of rationality including the capacity for morality (Hill 1992: 38–57, esp. 40–1) and others have maintained that only a more fully realized rational commitment to moral action qualifies someone as an end in herself (Dean 1996).
- 5 See also Kant's discussion of 'the state of health proper to a human being' (Kant 1996: 148; MM 6: 384), and Kant's claim that reason can only tell us of what exists or has 'reality,' not of negations (Kant 1998a: 554–5; CPR A574–6).
- 6 The phrase 'nicht der Mensch' appears in the first edition of the book, with which Kant was most directly involved, and in the Akademie edition, but is omitted from the second edition and some translations.
- 7 See also Kant 1987: 74; C3 5: 227.
- 8 For more on Kant's specific criticism of Leibniz and Wolff's perfectionism, see Kant 2002: 243; G 4: 443; Kant 1997a: 36; C2 5: 40–1.
- 9 This also fits with Kant's statement in his general introduction to the discussion of duties to others, where he says that 'one can love one's neighbor though he might deserve but little respect' (Kant 1996: 198; MM 6: 448).
- 10 For helpful comments or inspiration I thank James O'Shea, Stephen Engstrom, two referees for *Kantian Review*, the participants in the workshop *Revisiting Issues in Moral and Political Philosophy*, held in Chapel Hill NC in May 2011, and especially the late Richard Smyth.
- 11 See especially Dean 2006, but also some related papers.
- 12 Denis does not have in mind here the position developed in this paper, that humanity is best taken to be an idea of reason (I had not proposed that yet). Nevertheless, the passages do support taking humanity as an ideally rational or good will.

References

- Dean, Richard (2006) *The Value of Humanity in Kant's Moral Theory*. Oxford: Oxford University Press.
- Denis, Lara (2010) 'Humanity, Obligation, and the Good Will: An Argument Against Dean's Interpretation of Humanity'. *Kantian Review*, 15, 118–41.

- Hill, Thomas E Jr (1992) *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca, NY: Cornell University Press.
- Kant, Immanuel (1987) *Critique of Judgment*, trans. Werner Pluhar. Indianapolis: Hackett Publishing Co.
- (1996) *The Metaphysics of Morals*, trans. Mary Gregor. Cambridge: Cambridge University Press.
- (1997a) *Critique of Practical Reason*, trans. Mary Gregor. Cambridge: Cambridge University Press.
- (1997b) *Lectures on Ethics*, trans. Peter Heath. Cambridge: Cambridge University Press.
- (1998a) *Critique of Pure Reason*, ed. Paul Guyer and Allen Wood. Cambridge: Cambridge University Press.
- (1998b) *Religion Within the Limits of Reason Alone*, ed. Allen Wood and George di Giovanni. Cambridge: Cambridge University Press.
- (2002) *Groundwork for the Metaphysics of Morals*, ed. Thomas E. Hill, Jr. and Arnulf Zweig. Oxford: Oxford University Press.
- (2007) *Lectures on Pedagogy*, trans. Robert Louden, in *Anthropology, History, and Education*, ed. Günter Zöller and Robert Louden. Cambridge: Cambridge University Press.
- Korsgaard, Christine (1996) *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- Leibniz, Gottfried (1969) *Monadology*. In *Philosophical Papers and Letters*, trans. Leroy Loemker, Dordrecht: Reidel.
- Wolff, Christian (1976) *Vernünfftige Gedancken von der Menschen Tun und Lassen zu Beförderung ihrer Glückseligkeit*. Hildesheim: Georg Olms.
- (1983) *Vernünfftige Gedancken von Gott, der Welt und der Seele des Menschen, auch alle Dingen Überhaupt*. Hildesheim: Georg Olms.
- Wood, Allen (1999) *Kant's Ethical Thought*. Cambridge: Cambridge University Press.