

Reconsidering RGV, AA 06:26n and the meaning of ‘Humanity’

At RGV, AA 06:26n Kant famously (or infamously) claims that humanity and personality are not necessarily coextensional. This claim has been characterized in the secondary literature as Kant’s worst mistake and as an unnecessary repudiation of his earlier (and more plausible) ethical thought. I argue that this characterization of RGV, AA 06:26n rests on a misinterpretation of the term ‘humanity’. I try to show that Kant’s claim at RGV, AA 06:26n not only is not problematic; it constitutes a powerful reminder of the kind of epistemic modesty that Kant argues for in the *Critique of Pure Reason*.

In part I of the *Religion within the Boundaries of mere Reason*, Kant makes the following claim:

We cannot consider this predisposition [*viz.*, the predisposition to personality] as already included in the concept of the preceding one [*viz.*, the predisposition to humanity], but must treat it necessarily as a special predisposition. For it certainly does not follow from the fact that a being has reason that this reason contains a capacity to determine the power of choice unconditionally through the bare representation of the qualification of its maxims for universal lawgiving, and thus [that this reason contains a capacity] to be practical for itself; at least, not so far as we can see. (RGV, AA 06:26n)¹

This short passage has acquired some notoriety. It has been characterized as Kant’s worst mistake; as constitutive of an unnecessary capitulation to alternative schools of thought; and as a repudiation of Kant’s earlier (and more philosophically plausible) ethical theory.

In this paper I argue that these characterizations of RGV, AA 06:26n are mischaracterizations. I argue that RGV, AA 06:26n represents a confirmation of one of the central tenets of Kant’s moral philosophy. That is, I argue that RGV, AA 06:26n gives evidence

¹ All citations are according to Akademie Edition pagination. All translations are my own. The German for this passage runs as follows: “*Man kann diese nicht, als schon in dem Begriff der vorigen enthalten, sondern man muß sie notwendig als eine besondere Anlage betrachten. Denn es folgt daraus, daß ein Wesen Vernunft hat, gar nicht, daß diese ein Vermögen enthalte, die Willkür unbedingt, durch die bloße Vorstellung der Qualifikation ihrer Maximen zur allgemeinen Gesetzgebung zu bestimmen, und also für sich selbst praktisch zu sein; wenigstens so viel wir einsehen können.*”

that, even by the time of the *Religion within the Boundaries of mere Reason*, Kant has not given up the belief that, although it is rationally impossible for me to believe I am not free, it is nonetheless logically possible that I am not free.

This paper is divided into two sections. In the first section, I reproduce one of the arguments from the secondary literature that attempts to show that RGV, AA 06:26n is inconsistent with Kant's earlier ethical theory. In the second section, I show why this argument does not work and give evidence for my reading of RGV, AA 06:26n.

I. Why some commentators think that RGV, AA 06:26n is a mistake

According to the standard reading of RGV, AA 06:26n, the issue being confronted in this passage is whether reason can be practical. If this is correct, then RGV, AA 06:26n constitutes an unwarranted repudiation of one of Kant's signature theses, *viz.*, that reason can be practical. According to this reading, RGV, AA 06:26n is nothing less than Kant telling us that the edifice of his practical philosophy is built on an insecure foundation. The argument for this reading is as follows.²

The claim at RGV, AA 06:26n is that "so far as we can see" there could be beings with the predisposition to humanity but without the predisposition to personality. I shall follow the conventions in the secondary literature and not distinguish between the predisposition to humanity and humanity or between the predisposition to personality and personality. So the claim in RGV, AA 06:26n is that 'humanity' and 'personality' are not necessarily coextensional.

² The argument reproduced here is taken from Wood, Allen: "Humanity as an End in Itself". In Paul Guyer (ed.), *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*, 165-189. Lanham, Boulder, New York and Oxford. 1998, 172. See Wood, Allen: *Kant's Ethical Thought*. Cambridge, England. 1999, 364n1 for a similar argument. Similar reasoning also seems to underlie the remarks in Korsgaard, Christine: *Creating the Kingdom of Ends*. Cambridge, England. 1996, 110-114.

To make any sense of this claim, one must know what Kant means by ‘humanity’. At TL, AA 06:441, Kant distinguishes between humanity and the person (he tells us that this allows us to contemplate even a rogue with pleasure). This indicates that humanity is a capacity or property of a person and not the person, being or creature itself. This is supported by the Formula of Humanity (FH), where Kant says that humanity is something “in” persons (GMS, AA 04:428).

At TL, AA 06:392 and TL, AA 06:423, humanity is contrasted with animality, indicating that humanity is one of our “higher” faculties or capacities. Both of these passages in the text indicate that humanity involves reason in some way.

At TL, AA 06:463 (and also at TL, AA 06:392) Kant claims that the distinguishing feature of humanity is the power to set ends and that one should respect it even in those who make themselves unworthy of it. This is in keeping with the last few passages. The power to set ends is a capacity involving reason.

In his *Lectures on Anthropology*, Kant divides the predisposition to humanity into a technical predisposition (which includes all of one’s learned skills and deliberative abilities designed for use toward arbitrary ends) and a pragmatic predisposition (which is the basis for one’s ability to compare one’s contingent ends and organize them into a systematic whole — viz., happiness) (Anth, AA 07:322-324; cf. KpV, AA 05:426-427).³ Again, Kant seems to be emphasizing the capacity to set ends (or things involved with this capacity) as being picked out by ‘humanity’.

Does ‘humanity’ pick out more than the power to set ends? At GMS, AA 04:430, in his discussion of the duty of self-cultivation, Kant seems to include our capacities for greater perfection and our fortunate natural aptitudes (the natural aptitudes that any particular human has, which will vary from human to human) in humanity. However, in the analogous discussion in the *Metaphysics of Morals* humanity is simply the power to set ends. In the *Metaphysics of Morals*, development of talents and cultivation of the self are not instances of promoting

³ Cf. Wood, Allen: “Humanity as an End in Itself”. In Paul Guyer (ed.), *Kant’s Groundwork of the Metaphysics of Morals: Critical Essays*, 165-189. Lanham, Boulder, New York and Oxford. 1998, 172; Wood, Allen: “Kant on Duties Regarding Nonrational Nature: P”, *Proceedings of the Aristotelian Society*, Supplementary Volumes, Vol. 72 (1998), 189-210, 189; Wood, Allen: *Kant’s Ethical Thought*. Cambridge, England. 1999, 119; and Wood, Allen: *Kantian Ethics*. Cambridge, England. 2008, 88.

humanity, but rather that which is required in order to make one worthy of humanity (TL, AA 06:392).

Comparing the passage in the *Groundwork for a Metaphysics of Morals* with the passage in the *Metaphysics of Morals* might lead one to conclude that Kant does not use ‘humanity’ univocally throughout his corpus. But there does seem to be a unifying theme; humanity picks out the power to set ends, or the rational capacity to set and organize ends.⁴

This gets us halfway through the claim at RGV, AA 06:26n. In order to make sense of the rest of the claim, one must know what Kant means by ‘personality’. This term is not used as frequently as ‘humanity’. Kant seems to think that ‘personality’ refers to the capacity that accounts for someone’s being rational and responsible (RGV, AA 06:26). He defines ‘personality’ explicitly as “the susceptibility to respect for the moral law as of itself sufficient incentive to the power of choice” (RGV, AA 06:27).⁵

With this understanding of ‘humanity’ and ‘personality’, one can return to examine the original claim from RGV, AA 06:26n. The claim was that ‘humanity’ and ‘personality’ are not necessarily coextensional. According to the standard reading, to say that ‘humanity’ and ‘personality’ are not coextensional is to say that there could be beings that have the capacity to set ends but that are not susceptible to the moral law as an of itself sufficient incentive to the power of choice. In other words, according to the standard reading, RGV, AA 06:26n is telling us that there could be beings that have the capacity to set ends but that are not subject to the moral law. Why might it seem implausible to say that ‘humanity’ and ‘personality’ are not coextensional? What is the problem with RGV, AA 06:26n on this reading?

⁴ However, Rawls argues that ‘humanity’ in FH does not include any reference to the general capacity to set ends (Rawls, John: *Lectures on the History of Moral Philosophy*. Edited by Barbara Herman. Cambridge, Massachusetts. 2003, 188). Guyer points out that this omission has significant consequences for Rawls’s interpretation of Kant, requiring Rawls to introduce the notion of “basic” or “true” human needs in order to generate positive duties and making it difficult for Rawls to explain the relation between morality and happiness that Kant thinks follows from his principle more generally (Guyer, Paul: *Kant*. USA and Canada. 2006, 395n8). If the exegesis in this paper is correct, then ‘humanity’ in the *Groundwork for a Metaphysics of Morals* does involve the general power to set ends (and perhaps other things, such as our natural aptitudes, too).

⁵ The original text runs as follows: “*die Empfänglichkeit der Achtung für das moralische Gesetz, als einer für sich hinreichenden Triebfeder der Willkür*”.

The rational capacity to set and organize ends involves the capacity to make comparative judgments of value. And the capacity to make comparative judgments of value involves an awareness of standards of value. But if Kant is correct, morality is the foundation of all such standards. Therefore, any being that has the rational capacity to set and organize ends will have the capacity to recognize and respond to the incentives of morality, thus will have personality.

Kant's seeming denial of this at RGV, AA 06:26n looks like an admission that it would be possible for there to be beings with the power to set ends but who are not susceptible to the incentives of morality. That is, this reading pushes us into saying that at RGV, AA 06:26n Kant is admitting the possibility of beings that are negatively free but that nonetheless are not subject to the moral law. This would be disastrous for Kant's philosophy. Moreover, it looks like a recantation of a (philosophically plausible) argument that Kant makes in the opening pages of part III of the *Groundwork for a Metaphysics of Morals*, where he asserts the connection between negative and positive freedom.

So (the argument concludes) RGV, AA 06:26n should be overlooked as anomalous in charitable reconstructions of Kant's philosophy. This, anyway, is the argument made by the standard reading of RGV, AA 06:26n. In the remainder of this paper, I shall argue that this reading of RGV, AA 06:26n is unwarranted. I shall propose a new reading of RGV, AA 06:26n that makes it compatible with the central tenets of Kant's practical philosophy.

II. A new reading of RGV, AA 06:26n

In the previous section, I reconstructed the standard reading of RGV, AA 06:26n. According to this reading, RGV, AA 06:26n represents a recantation of one of the central tenets of Kant's practical philosophy, *viz.*, that reason can be practical in itself. The argument for this reading can be summarized as follows. 'Humanity' picks out the capacity to set ends. 'Personality' picks out the susceptibility to the moral law. Therefore, in claiming that there could be beings

with humanity but without personality, Kant is claiming that there could be beings with the capacity to set ends that are not subject to the moral law. This is philosophically implausible and at variance with Kant's earlier work, hence we should overlook RGV, AA 06:26n in reconstructing Kant's philosophy.

In this section, I shall argue that this is a misreading of RGV, AA 06:26n. I shall argue that in the *Religion within the Boundaries of mere Reason*, 'humanity' is not used to pick out the capacity to set ends.⁶ Moreover, once we see how 'humanity' is used in the *Religion within the Boundaries of mere Reason*, RGV, AA 06:26n can be read as in accordance with Kant's practical philosophy.

I argued above that Kant does not use the term 'humanity' univocally throughout his corpus. As already pointed out, in the *Groundwork for a Metaphysics of Morals* 'humanity' seems to pick out the capacity to set ends and, in addition, an agent's natural aptitudes. By way of contrast, in the *Metaphysics of Morals*, Kant seems to use 'humanity' to pick out merely the capacity to set ends.

One might become more suspicious upon noticing that Kant uses 'personality' and 'humanity' interchangeably in the *Critique of Practical Reason* (KpV, AA 05:87). He does this also at some places in the *Metaphysics of Morals* (see, e.g., TL, AA 06:422-423). It is tempting to argue on these grounds that humanity is the capacity that, when exercised correctly, becomes personality; humanity is the will (the capacity to set ends according to reason) and personality is a good will (the actual setting of ends completely according to reason).⁷ But this could prove difficult to reconcile with the articulation of FH in the *Groundwork for a Metaphysics of Morals*, where Kant tells us that we ought to treat humanity as an end in persons. If one thinks that persons are those beings that have personality, then FH tells us to treat humanity as an end

⁶ One commentator who might support this thesis is Guyer, who argues that the humanity of the *Groundwork for a Metaphysics of Morals* includes both the humanity and the personality of the *Religion within the Boundaries of mere Reason* (Guyer, Paul: *Kant on Freedom, Law, and Happiness*. Cambridge, England. 2000, 192n16). Another commentator who might support this thesis is Savage. But Savage's interpretation of the text is slightly different from Guyer's. Savage, like Rawls (see above), takes the humanity of the *Groundwork for a Metaphysics of Morals* to pick out merely the personality of the *Religion within the Boundaries of mere Reason* (Savage, Denis: 'Kant's Rejection of Divine Revelation and his Theory of Radical Evil'. In Philip J. Rossi and Michael Wreen (ed.'s), *Kant's Philosophy of Religion Reconsidered*, 54-77. Bloomington and Indianapolis, 1991, 75n2).

⁷ This claim can be found in Korsgaard, Christine: *Creating the Kingdom of Ends*. Cambridge, England. 1996, 114: "Humanity, completed and perfected, becomes personality".

in beings with personality. So how could humanity become personality once completed and perfected? What meaning would FH take on if humanity, when exercised correctly, became personality? Not an intuitive one, I think. More to the point, the fact that Kant uses ‘humanity’ and ‘personality’ interchangeably in the *Metaphysics of Morals* and in the *Critique of Practical Reason* but goes on in the *Religion within the Boundaries of mere Reason* to claim that these terms are not extensionally equivalent (let alone intensionally equivalent) indicates that he is using both of these terms differently in these different texts.

The definition of ‘personality’ as the susceptibility to the moral law is taken from the *Religion within the Boundaries of mere Reason*. But the standard definition of ‘humanity’ as the power to set ends was culled from a variety of places in Kant’s work, none of which was from the *Religion within the Boundaries of mere Reason*. In order to see whether the standard reading of RGV, AA 06:26n is correct, I shall look at how Kant uses ‘humanity’ in the *Religion within the Boundaries of mere Reason*.

In the passage to which the footnote at RGV, AA 06:26n is appended, Kant explains the predisposition to humanity as “self-love which is physical and yet involves comparison” (RGV, AA 6:27)⁸ and he makes the following remark:

The first [animality] has no [kind of] reason at its root, the second [humanity] does have practical reason at its root, but [a kind of practical reason that] *is only subservient to other incentives*, but the third [personality] alone has at its root [a kind of] reason that is practical of itself, that is, reason that gives laws unconditionally. (RGV, AA 06:28)⁹

This suggests that the humanity of the *Religion within the Boundaries of mere Reason* does not involve the general power to set ends like the humanity of the *Groundwork for a Metaphysics of Morals* and the *Metaphysics of Morals*. Rather, humanity in the *Religion within the Boundaries of mere Reason* involves reason subservient to other incentives; this suggests that reason in humanity in the *Religion within the Boundaries of mere Reason* is merely instrumental. Rather than being involved in the choosing and setting of ends, humanity in the *Religion within the*

⁸ The original text runs as follows: “*der zwar physischen, aber doch vergleichenden Selbstliebe*”.

⁹ My emphasis. The original text runs as follows: “*die erste keine Vernunft, die zweite zwar praktische, aber nur andern Triebfedern dienstbare, die dritte aber allein für sich selbst praktische, d.i. unbedingt gesetzgebende Vernunft zur Wurzel habe*”.

Boundaries of mere Reason is involved only in choosing and setting means to achieving other ends that we have as a result of our desires and inclinations. Reason in humanity is completely subservient to other incentives; it is not practical in itself.

It might be the case that purely instrumental rationality of this sort is not logically possible and that any being that can be instrumentally rational *eo ipso* is responsive to incentives of pure reason (i.e., to the moral law). But this is contentious. And the argument from the standard reading is directed against a different foe. The argument from the standard reading is directed against the attempt to claim that a being can have the power to set ends (negative freedom; the capacity to choose ends and not be determined by inclination) without the power to behave morally (positive freedom; the capacity to legislate unconditional laws to oneself). As pointed out above, this is something that Kant argues for in part III of the *Groundwork for a Metaphysics of Morals*. In part III of the *Groundwork for a Metaphysics of Morals* Kant himself argues that if a being has the power to set ends (negative freedom), it has the power to behave morally (positive freedom). But it is not clear that this argument has any bearing on whether the humanity of the *Religion within the Boundaries of mere Reason* and the personality of the *Religion within the Boundaries of mere Reason* are coextensional.

In other words, the argument above seems like good philosophy; it seems plausible that negative freedom implies positive freedom. Indeed, if Kant were using ‘humanity’ in the *Religion within the Boundaries of mere Reason* to refer to the capacity to set ends, then RGV, AA 06:26n does not make sense and goes against Kant’s own argument in section III of the *Groundwork for a Metaphysics of Morals*. Maybe the concept of the capacity to set ends is not contained and does not itself contain the concept of personality. But a being with the capacity to set ends is free in the negative sense, thus free in the positive sense, thus has a reason capable of being practical, and thus has personality regardless of the concept containment relations. And Kant believed this in the *Groundwork for a Metaphysics of Morals*. But at RGV, AA 06:26n Kant is using ‘humanity’ in a different way than he does in the *Groundwork for a Metaphysics of Morals*. Thus, the argument above does not seem to show that what Kant says at RGV, AA 06:26n about the possibility of a being with humanity but without personality is false.

In point of fact, if one follows this line of thought (i.e., if one agrees that Kant uses ‘humanity’ in the *Religion within the Boundaries of mere Reason* to refer to instrumental reason rather than to the power to set ends), then RGV, AA 06:26n seems to be a powerful reminder of a central tenet of Kant’s moral philosophy. A being with purely instrumental reason does not have the capacity to set ends, hence is not negatively free (thus no positive freedom and no reason qua practical faculty. A being with only instrumental reason would not have a will in the Kantian sense).¹⁰ What is RGV, AA 06:26n telling us on this reading?

There is a distinction between what is logically necessary and what is rationally necessary. It is rationally necessary that, if I believe A and if A then B, I not believe not B. However, it is not logically necessary that if I believe A and if A then B, I not believe not B. I could believe A and if A then B and believe not B. It would be strange — irrational — but not logically impossible.

One way of putting Kant’s position with regard to freedom is that it is rationally necessary that I not believe that I am not free.¹¹ However, it is not logically necessary. It is possible for me to believe that I am not free. It would be irrational, but it is possible. And this says nothing about whether I actually am free. It is rationally necessary that I not believe that I am not free, but it is logically possible for me to believe that I am not, and it is logically possible that, regardless of what I believe, I am not free.

Continuing with this line of thought, it is, according to Kant, rationally necessary that I not believe that I do not have the capacity to set ends. It is not logically necessary that I not believe this, and it is logically possible that I do not have this capacity. But it is rationally necessary that I not believe that I do not.

What RGV, AA 06:26n is telling us is that as far as we can see,¹² it is logically possible that there are beings with purely instrumental reason. Maybe you and I are such beings. It is

¹⁰ Another way to see this is to appeal to the fact that it is personality that brings in imputability at RGV, AA 06:26 (cf. MS, AA 06:223) and to note Kant’s remarks about the close relationship between freedom and imputability (cf. R7129 and R7130).

¹¹ The length constraints on this paper do not allow me to look at why Kant believes this.

¹² “... *wenigstens so viel wir einsehen koennen*” (RGV, AA 06:26n).

logically possible that what we take to be setting ends is some sort of deterministic turnspit; perhaps we do not have the capacity to set ends (humanity in the sense of the *Groundwork for a Metaphysics of Morals*); perhaps we are mere Humean beings — it is logically possible — and thus have neither the humanity of the *Groundwork for a Metaphysics of Morals* nor the personality of the *Religion within the Boundaries of mere Reason*.

Of course, none of this is rationally possible. But RGV, AA 06:26n is cautioning us against confusing rational necessity with logical necessity. RGV, AA 06:26n is making an epistemological point that recalls us to the epistemic modesty argued for in the *Critique of Pure Reason*. RGV, AA 06:26n is telling us that although we ought to believe that we are free, it is possible for us not to believe this and, more to the point, it is possible that we are not.