

This volume of new essays provides a comprehensive and structured examination of Kant's justification of norms, a crucial but neglected theme in Kantian practical philosophy. The essays engage with the view that a successful account of justification of normative claims has to be non-metaphysical and go on to pursue further implications in ethics, legal and political philosophy, and philosophy of religion.

### Contributors

Henry E. Allison  
Karl Ameriks  
Sorin Baiasu  
Paul Guyer  
John Hare  
Otfried Höffe  
Larry Krasnoff  
Adrian W. Moore  
Andrews Reath  
Sebastian Rödl  
Houston Smit  
Robert Stern  
Mark Timmons  
Howard Williams  
Allen Wood

**OXFORD**  
UNIVERSITY PRESS

[www.oup.com](http://www.oup.com)

Cover image: ©PhotoSky 4t.com / Shutterstock.com

ISBN 978-0-19-539568-6



9 780195 395686

TIMMONS

BAIASU

KANT ON PRACTICAL JUSTIFICATION

OXFORD



**KANT ON PRACTICAL JUSTIFICATION**

**MARK TIMMONS AND SORIN BAIASU**

- Guyer, P. (2008) *Knowledge, Reason, and Taste: Kant's Response to Hume*. Princeton, NJ: Princeton University Press.
- Kant, I. (1996a) *Groundwork of The Metaphysics of Morals*, in *Practical Philosophy*, 41–108. Tr. and ed. M. J. Gregor. Cambridge: Cambridge University Press.
- Kant, I. (1996b) *Critique of Practical Reason*, in *Practical Philosophy*, 137–271. Tr. and ed. M. J. Gregor. Cambridge: Cambridge University Press.
- Korsgaard, C. M. (1999) "Self-constitution in Plato and Kant," *Journal of Ethics* 3(1): 1–29. Reprinted in *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press (2008).
- Korsgaard, C. M. (2009) *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Quarfoot, M. (2006) "The Circle and the Two Standpoints," in C. Horn and D. Schönecker (eds.), *Groundwork for the Metaphysics of Morals*, 285–300. Berlin: Walter de Gruyter.
- Rawls, J. (1975) "A Kantian Conception of Equality," *Cambridge Review* 96: 94–9, reprinted in Rawls (1999).
- Rawls, J. (1980) "Kantian Constructivism in Moral Theory," *Journal of Philosophy*. 77(9): 515–72.
- Rawls, J. (1993) *Political Liberalism*. New York: Columbia University Press.
- Rawls, J. (1999) *Collected Papers*. Ed. S. Freeman. Cambridge, MA: Harvard University Press.
- Rawls, J. (2000) *Lectures on the History of Moral Philosophy*. Ed. B. Herman. Cambridge, MA: Harvard University Press.
- Schönecker, D. (1999) *Kant: Grundlegung III: Die Deduktion des kategorischen Imperativs*. Freiburg and Munich: Verlag Karl Alber.
- Stern, R. (ed.) (1999) *Transcendental Arguments: Problems and Prospects*. Oxford: Clarendon Press.
- Strawson, P. F. (1966) *The Bounds of Sense*. London: Methuen.
- Timmons, M. (ed.) (2002) *Kant's Metaphysics of Morals: Interpretative Essays*. Oxford: Oxford University Press.
- Walker, R. C. S. (2006) "Kant and Transcendental Arguments," in P. Guyer (ed.), *The Cambridge Companion to Kant and Modern Philosophy*, 238–68. Cambridge: Cambridge University Press.
- Wallace, R. J. (1994) *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.
- Wood, A. W. (2006) "The Supreme Principle of Morality," in P. Guyer (ed.), *The Cambridge Companion to Kant and Modern Philosophy*, 342–80. Cambridge: Cambridge University Press.

{ 9 }

## Formal Approaches to Kant's Formula of Humanity

Andrews Reath

My aim in this chapter is to explore different ways of understanding Kant's Formula of Humanity (FH) as a formal principle. I believe that a formal principle for Kant is a principle that is constitutive of some domain of cognition or rational activity. It is a principle that both constitutively guides that activity and serves as its internal regulative norm. In the first section of this chapter, I explain why it is desirable to find a way to understand the Formula of Humanity as a formal principle in this sense. In sections II and III I discuss two interpretive approaches to Kant's idea that rational nature or humanity is an end in itself, both of which may be construed as treating the Formula of Humanity as a formal principle. By focusing on the notion of formal principle, I hope to raise a set of issues about how to understand the idea of rational nature or humanity as an end in itself, and about the relation of the Formula of Humanity to the Formula of Universal Law (FUL). I do not resolve the issues in this chapter, though I briefly sketch some resolution at the end.

### I. Two Poles to the Formula of Humanity

In the first section, I describe what we might think of as two poles of thought about FH, and then lay out a partial list of desiderata for an interpretation of FH.

#### 1. RATIONAL NATURE AS THE SUBSTANTIVE VALUE OF MORAL THOUGHT

The idea of rational nature or humanity as an end in itself introduces, or perhaps just makes explicit, the end that serves as the substantive value that animates moral thought and concern. Kant argues that the existence of an end in

itself—an end of absolute and incomparable worth—is both a sufficient and a necessary condition of an authoritative categorical imperative and that persons are such ends: “rational beings are called *persons* because their nature already marks them out as an end in itself, that is as something that may not be used merely as a means, and hence so far limits all choice (and is an object of respect)”<sup>1</sup> (*GMS* 4: 428). The “absolute worth” of rational nature (and the ensuing principle of humanity as an end) is the “ground of determinate laws” that specify the proper attitude toward and treatment of persons and serves as the “supreme limiting condition” of all free action and choice of subjective ends (*GMS* 4: 429, 431). Arguably the idea of rational nature as an end in itself would provide Kant’s answer to questions about the “subject matter of morality.”<sup>2</sup> Morality is about persons—more specifically, it is about acting from principles that show proper respect for persons as rational agents with autonomy.

As ordinarily understood, the principle of respecting persons as ends in themselves is a requirement to acknowledge the moral standing that gives persons claims to certain kinds of consideration and treatment. It directs our attention as reasoners and agents to our attitudes toward persons (as rational agents) and to the ways in which our choices affect the rational capacities and interests of persons—as we might say, to persons, including ourselves, on the receiving end of our attitudes and actions. Let’s call this the requirement of “respect for persons (as rational agents) in the standard intuitive sense.” To many people, it is a way of framing what is at issue in moral concern that is both accurate and deeply appealing and that seems able to support the deliberative priority of moral considerations. One common philosophical rendition of respect for persons that aligns it with Kant’s FUL is that it is the requirement to act from principles that can justify one’s actions to those affected by them (as rational agents with autonomy), where the standard of justification is what can be willed as universal law for rational agents with autonomy. This understanding of Kant’s principle is supported by his remark in the deceptive promise example that to value rational beings as ends in themselves is to value them as “beings who must be able to contain in themselves the end of the very same action” (*GMS* 4: 429–30). One treats others as ends when they can rationally endorse one’s underlying maxim of action, where what is rationally endorsable is what all agents can jointly will as universal law (for agents with autonomy).<sup>3</sup>

<sup>1</sup> I use the translations of Kant’s *Groundwork*, *Critique of Practical Reason*, and *The Metaphysics of Morals* from Kant (1996).

<sup>2</sup> To use T. M. Scanlon’s phrase—see Scanlon (1998: 1–5).

<sup>3</sup> Here I follow Rawls (2000), “Kant III,” especially pp. 190–92. Rawls interprets Kant’s language that the recipient (here the promisee) be able to “contain in himself the end” [*den Zweck in sich enthalte*] of the action in terms of the recipient rationally endorsing the agent’s maxim by seeing that it can be willed as universal law. Maxims that can be willed as universal law are mutually endorsable, thus serve as principles through which we can justify our actions to each other. For other treatments of FH that align it with FUL, see Hill (1992: 45) and O’Neill (1989: 137–43).

The introduction of FH follows a question: “is it a necessary law for all rational beings always to appraise their actions in accordance with such maxims as they themselves could will to serve as universal laws?” (*GMS* 4: 426) The FUL is already in hand at this point in the *Groundwork* (both as the principle presupposed by the ordinary notion of duty and as the principle that expresses the form of a practical law or unconditional practical requirement, which would appear to be part of the very idea of practical reason). But its authority remains an open question that FH helps to address in (I suggest) two ways. First, FUL is an abstract practical principle whose authority it is natural to question. FH tells us that acting from FUL is about respecting persons and relations of mutual respect between persons. These are values with strong intuitive appeal that it makes sense to care about and that can command our allegiance. Seeing that respect for persons is at issue in FUL helps to deflect questions about its authority and to motivate acceptance. In this respect FH (along with the Formula of Autonomy and the idea of the realm of ends) brings the moral law (“an idea of reason”) “closer to intuition . . . and thereby to feeling” and “provide[s] access for the moral law” (*GMS* 4: 436, 437).<sup>4</sup>

Second, the sequence of the formulas of the Categorical Imperative (CI) through *Groundwork* II is part of a technical philosophical argument for the authority of the moral law. One component of this argument is the analytic claim that “a free will and a will under moral laws are one and the same,” which I understand as the claim the moral law is the formal (or internal) principle of free volition (*GMS* 4: 447). The sequence of reformulations of the CI sets up this claim by showing that the FUL can be understood as a principle of autonomy—as the principle through which the will is a law to itself—and accordingly is the formal principle of free volition. The introduction of FH must advance this argument in some way—presumably by contributing to the transition from subjection to duty to autonomy.<sup>5</sup>

To summarize this pole: rational nature as an end in itself is the substantive value that underlies moral thought and concern and it leads to a requirement of respect for persons. Its introduction advances the argument for the authority of the moral law in different ways. It “provides access” for the moral law since it is a value that it makes sense to care about, and it is a reformulation of the basic principle that contributes to the technical argument of the *Groundwork* as a whole.

<sup>4</sup> Even if the introduction of FH “provides access” for the moral law, it cannot close the question of its authority at this point. The question how categorical imperatives are possible arises initially because of the unconditional character of duty: how can there be requirements that are independent of and take priority over desire-based reasons? (Do we have the motivational capacities to act from such principles and does it make sense to accept their authority?) Since FH purports to be an unconditional requirement, it cannot by itself resolve that issue.

<sup>5</sup> For discussion of this transition, see Reath (2006: 99–108, 137–49) and Engstrom (2009: 135–6, 149–51).

## 2. FH AS A FORMAL PRINCIPLE

A different pole comes to the fore when we consider Kant's view that the various formulas of the CI are equivalent. Kant holds that the FUL is a formal principle. Among other things, he refers to it as the "formal principle of volition" (*GMS* 4: 400) and as the "formal practical principle of pure reason" (*KpV* 5: 41).<sup>6</sup> Indeed the FUL is both the formal principle of morality and, given the arguments of the opening of *Groundwork* III (*GMS* 4: 446–7), the formal principle of free rational agency. If FUL and FH are "at bottom only so many formulas of the very same law," (*GMS* 4: 436) then FH is likewise a formal principle that functions in these capacities—both as the formal principle of morality and as the formal principle of free rational agency. Obviously if one rejects Kant's claims about the equivalence of the formulas, there is no need to consider whether FH can be understood in this way. But the assumption that FH is *not* a formal principle is one barrier to accepting Kant's claim about equivalence. Since showing that FH can be understood as a formal principle would accordingly remove some skepticism about this claim, it is worth giving this approach to FH a hearing.

To see what this approach to FH involves, let me explain what I think formal principles are for Kant.<sup>7</sup> Kant tends to regard the fundamental principles in some domain of cognition or rational activity as formal principles. This is quite clear in his moral philosophy, where he is explicit that the fundamental principle of morality must be a formal principle and that only a formal principle—a principle that determines the will through its form rather than its matter and that prescribes the formal condition of universal law—has the necessity of a practical law.<sup>8</sup> The connection between form and normative necessity is explained if we understand a formal principle as the internal constitutive principle of a domain of cognition or rational activity. It is the principle that defines or describes and makes it possible to engage in that activity, thus the principle that any subject engaged in that activity must follow. So understood, the formal principle of a domain of cognitive activity is uniquely suited to govern it with normative necessity because it is not coherently rejected by anyone engaged in that activity.

The contemporary conception of formal principles focuses on abstraction from content, as do many of Kant's own discussions of the basic principle of morality. As we know, the normative force of a categorical imperative or practical law does not depend on any purpose or an interest in the matter of the principle, but only on its form, and to that extent such a principle "abstracts

<sup>6</sup> Cf. *GMS* 4: 400: that in action done from duty the will "is determined by the formal principle of volition as such"; and *GMS* 4: 444: a good will, "whose principle must be a categorical imperative... contains merely the form of volition as such and indeed as autonomy..." indicating that the CI captures the form of volition as such.

<sup>7</sup> This and the next paragraph draw on section III of Reath (2010).

<sup>8</sup> Cf. *KpV* 5: 27, 34, 39, 41, 64.

from all objects."<sup>9</sup> But the fact that the normative force of a practical law is independent of its matter provides no insight as to why only a formal principle can serve as a practical law, and does not explain what it means for its authority to depend on its form. Understanding formal principles as internal constitutive principles provides a positive explanation of their normative authority and foundational role.

A formal principle of a domain of cognition grows out of and expresses the self-understanding of that activity. It would appear that any kind of rational activity understands itself as having certain features that make it what it is—indeed that it is a formal feature of rational activities that they understand themselves to have a certain form—and that all genuine instances of the activity are normatively guided by this self-understanding. (Rational activity is self-conscious and is guided by its awareness of what it is.) The spontaneity of cognition or rational activity, in part, is that it is normatively guided by this self-understanding (of its own form).

In order to make this idea a bit less abstract, let me illustrate with an example taken from Stephen Engstrom (much simplified).<sup>10</sup> Engstrom suggests that it is the mark of judgment that it is "self-consciously self-sustaining." The self-sustaining component is that a judgment understands itself to make an objectively valid claim that excludes incompatible claims and that agrees with all other judgments and is confirmed by this agreement. Judgment is *self-consciously* self-sustaining because it sustains itself through its understanding that it is making an objectively valid claim. Among other things that means that judgment self-consciously seeks agreement with all other judgments as its formal aim, both what Engstrom calls "subjective agreement" and "objective agreement." "Subjective agreement" is that all judging subjects are to agree with or hold a valid judgment, and "objective agreement" is that judgments with different content are to agree with and support each other. Thus the formal feature of judgment is that it understands itself, and so constitutively aims, to fit together with all other judgments in a single (mutually supporting) body of knowledge that holds for all judging subjects, and moreover that it sustains itself through its consciousness that it does fit together with all other judgments in this way. In the case of theoretical judgments of the understanding, since the categories and principles of the understanding are conditions of agreement or unity in one objective self-consciousness, they serve as the internal norms of judgment.

<sup>9</sup> See *GMS* 4: 441: A categorical imperative "must abstract from all objects to this extent [*von allem Gegenstande sofern abstrahieren*]; that they have no influence on the will, so that practical reason (the will) may not merely administer an interest not belonging to it, but may simply show its own commanding authority as supreme lawgiving." In other words, a categorical imperative "abstracts from all objects" in the sense that its normative force does not depend on an empirically given interest in some object. A categorical imperative must carry its authority in itself, because it has the very form of volition.

<sup>10</sup> Here I summarize Engstrom (2009: 98–118).

The structure here is that a judgment—in this case a theoretical judgment—understands itself to be making an objectively valid claim that stands with all other judgments in one body of knowledge. This is a necessary feature of judgment, in that a mental state that does not understand itself in this way is not a judgment. Further, this self-understanding leads to a set of internal principles that govern exercises of judgment in two respects. First, the internal principles describe and constitutively guide the operation of theoretical judgment and, because they are part of its self-understanding, tacitly guide all instances, even false judgments. One judges about items given in intuition (brings them to the objective unity of self-consciousness [*KrV* B141]) by bringing them under the categories and principles of the understanding. Second, these internal principles function as regulative norms that, again because they are based in the self-understanding of judgment, set authoritative standards of success and failure. A judgment that does not meet the condition of agreement with all other judgments must be withdrawn.<sup>11</sup>

What goes into construing FUL as a formal principle? For present purposes, I'd like to assume with Kant that FUL is the formal principle (the internal constitutive norm) of rational volition. FUL is then the principle that expresses the self-understanding that both (1) constitutively guides or describes rational volition, and (2) serves as its regulative norm. In its second capacity as regulative norm, it is familiar to us in imperatival form as the principle of a good will. In its first, constitutive capacity it is the internal principle that describes the operation of free rational volition. Since this notion is harder to come to terms with, let me say a bit about it, though without taking on the many large issues that it raises. Rational volition (I hold) understands itself to specify action by deriving actions from universal principles that provide sufficient rational support.<sup>12</sup> That is, rational choice understands itself as part of its form to aim at actions and ends supported by good and sufficient reasons, and it is guided by that self-understanding. This self-understanding is expressed by FUL—or so I interpret Kant's view. In that case FUL describes the operation of the will: volition involves deriving or specifying action through what are taken to be universally valid principles, or judgments of good reasons; and it tacitly guides all exercises of the will, including bad choice that does not conform to this principle in its regulative-normative capacity.

<sup>11</sup> Christine Korsgaard has also developed the idea that constitutive principles are both descriptive of an activity and normative. See, among other places, Korsgaard (2008: 7–10).

<sup>12</sup> In fact, following Engstrom, the self-understanding of rational volition is more complex. First, it understands itself both to be a form of practical thinking, that is a form of thought that can bring its object about, simply through its self-understanding as efficacious. This aspect of its self-understanding leads to norms of instrumental rationality (here see Engstrom 2009: 28–44). Second, it understands itself to aim at actions and ends judged to be good. I focus on the second aspect here.

One way to represent this conception of volition philosophically has been suggested by Barbara Herman. Kant tells us that the will is a capacity to derive actions from a representation of certain laws or principles (cf. *GMS* 4: 412). Herman adds that “among the laws that we can and do represent to ourselves is the law that is constitutive of the will's own causal power” (Herman 2007: 171). In all rational volition “an agent is moved by a perceived connection of the action to her representation of herself willing an end, which is to say, according to a representation of the will's constitutive principle... the principle constitutive of the will's own activity... [is] what we (always and necessarily) represent to ourselves in and as a condition of rational choice” (Herman 2007: 246). So here is a way to unpack the idea that rational volition is governed by its own self-understanding as aimed at good and sufficient reasons: all rational volition proceeds from a representation of the formal principle of volition and understands itself to specify action through the application of this principle—that is by deriving action from principles taken to be universally valid or to provide sufficient rational support. A bit more work will tell us that rational volition, so conceived, is free activity. It is robustly self-determining because it is governed by its self-understanding of its form as expressed in its internal norm, independently of certain kinds of outside influence.<sup>13</sup> Thus FUL is the formal principle of *free* agency.

Now the point I wish to make is that what we just said about FUL must also hold for and map onto FH in some way. FH must be the formal principle of morality, the principle of a good will. (So much is obvious. The issue here is whether it is actually the same norm as FUL.) At the same time it is the formal principle of free rational volition—the principle that describes the operation of rational volition and tacitly guides all instances, even those that fail to satisfy the moral norm. So, for example, if a representation of FUL in some way figures in all rational volition, so must a representation of rational nature as an end in itself.

<sup>13</sup> The idea is that this conception of rational volition satisfies Kant's conception of transcendental freedom. It is negatively free in various senses: it is governed normatively rather than causally, and since it is not bound to take its desires to indicate reasons, it is motivationally independent. Positively, it is guided by its own self-understanding as expressed in its formal principle—where that is a principle that it gives to itself a priori through its own self-understanding. Taking FUL to be the formal principle of free agency provides a nice account of how bad action is free: it is free because guided by a representation of the formal principle of the will, bad because that principle is *mis*represented (see Herman 2007: 171–2, 246). (On Engstrom's view the formal feature of all free action is that it contains the presupposition of universality; in morally good choice, the content of the maxim agrees with its form, while in morally bad choice it does not; see e.g., Engstrom 2009: 131–4.) If the thesis that FUL constitutively guides free volition is to succeed, it has to provide some substantive guidance—e.g., by setting out obligatory ends that can initiate practical reasoning. For recent accounts, see “The Scope of Moral Requirement,” section III, and “Obligatory Ends” in Herman (2007), and Engstrom (2009: 188–223).

## 3. SOME DESIDERATA FOR A READING OF FH

With these remarks in mind, let me suggest a partial list of desiderata for a reading of FH.

1. If one takes seriously Kant's claim that FUL and FH are equivalent, a reading of FH should show that it is both the formal principle of morality and the formal principle of free rational volition—and moreover a formulation of this principle that is recognizably equivalent to FUL.
2. It should preserve the idea that rational nature as an end in itself (and respect for persons in the “standard intuitive sense”) is the substantive value that underwrites moral thought and concern. I think that this is clearly Kant's intent and it is an important feature of his moral conception. Since, the equivalence of FUL and FH also presupposes that this substantive value is implicit in FUL, this point suggests a desideratum for our understanding of FUL. (Formal practical principles, in Kant's sense, need not be devoid of substantive value commitments.)
3. It should show how FH restates the moral law (as previously expressed by FUL) in a way that advances the argument for its authority, both (a) bringing the moral law closer to intuition and to feeling, and (b) advancing the overall philosophical project of the *Groundwork*. A reading of FH that satisfies desideratum 2 will satisfy 3a, since respect for persons is a value that it makes sense to care about (thus “provides access for the moral law”). Aside from that, we want an understanding of the role of FH in the sequence of formulas that shows how it advances the overall argument of the *Groundwork*.
4. Of various passages that a reading of FH should fit and make sense of, I'll mention three. First, prior to introducing FH Kant writes that “in [an end in itself], and in it alone, would lie the ground of a possible categorical, imperative, that is, of a practical law”<sup>14</sup> (*GMS* 4: 428; my italics). After claiming that persons are ends in themselves, he then says that without an end in itself “nothing of absolute worth would be found anywhere; but if all worth were conditional and therefore contingent, then no supreme practical principle for reason could be found anywhere” (*GMS* 4: 428). It is clear that the existence of an end in itself is a *sufficient* ground of practical laws: if there are such ends, then there are laws governing proper responses to them. But the claim that an end in itself provides a ground that is *necessary* for a categorical

<sup>14</sup> In a closely related passage Kant writes that while “the ground of all practical law-giving lies objectively in the rule and the form of universality . . . subjectively, however, it lies in the end” (*GMS* 4: 431)—raising the question in what sense an end in itself (“the subject of all ends”) is the “subjective ground of law-giving.” The three that I mention (in addition to the argument at *GMS* 4: 429) are obviously not the only relevant passages.

- imperative should puzzle us. Kant's arguments for the authority of the CI—either the argument of *Groundwork* III or the Fact of Reason in the second *Critique*—do not directly refer to an end of absolute value. The FUL is a principle that appears able to stand on its own, and it specifies objective ends. In what sense is an end of absolute value a *necessary* ground of a categorical imperative?
5. Second, we want an explanation of the promissory note at *GMS* 4: 429n to the effect that the thesis that humanity is an end in itself will be made good in *Groundwork*, III. The claim that “every other rational being also represents his existence in this way [as an end in itself] consequent on just the same rational ground that holds also for me” is advanced as a “postulate” whose grounds are supplied in the third section (*GMS* 4: 429). Presumably the warrant for this claim has to do with the necessity of acting under the idea of freedom—that beings who necessarily act under the idea of freedom on that basis necessarily represent their existence as ends in themselves. That would show that rational agents necessarily represent themselves as ends in themselves, since the necessity of acting under the idea of freedom is a necessary feature of rational agency (and not just human agency). If so, the question to address is why the necessity of acting under the idea of freedom is a basis for representing one's existence as an end in itself.
  6. A third important passage is the remark at *GMS* 4: 437: “Rational nature is distinguished from the rest of nature by this, that it sets itself an end [*dass sie ihr selbst einen Zweck setzt*].” In the balance of this paragraph, Kant says that the end that rational nature sets for itself is “the matter of every good will,” is “an independently existing [*selbständiger*] end” that must always be valued as an end, and is the “subject of all ends” (which is the subject of a good will). Given what follows, it is clear that the opening sentence is claiming that rational nature sets itself a *single* end, namely *itself* as end in itself. Arguably every organized creature has itself as its own end—that is, has the end of maintaining itself in its form. What would distinguish rational nature, then, is that it *sets* itself as an end for itself—viz., that it freely and spontaneously makes itself its own end. This may mean that it self-consciously understands itself to be its own end and that this self-understanding in some sense guides its choices or activity.<sup>15</sup>

<sup>15</sup> Compare *GMS* 4: 412, where Kant claims that what sets rational beings apart from the rest of nature is that their activity is guided by principles that they self-consciously represent to themselves. Because they are aware of representing these laws to themselves, such laws can guide their activity normatively. Likewise, at *GMS* 4: 437, Kant may mean that every organized creature has itself (its form) as an end, but that only rational nature self-consciously sets itself as its end (and thus is normatively guided by that end).

Now these are “desiderata”: it would be desirable to have a reading that satisfies these criteria, which is not to say that it is possible. But one might think that readings of FH can be assessed by how well they do against this, or a more complete list of desiderata. In the interpretative approaches to FH considered in the next two sections, I focus mainly on 1, 2, 4, and 6 (sadly, I don’t yet have much to say about 5<sup>16</sup>).

## II: One Formal Reading: The End for the Sake of which Other Things Have Value

I shall now outline two different ways of understanding FH as a formal principle, (focusing on its role as the formal principle of free rational agency), and then consider how well each does by the above desiderata. The first interprets FH along the following lines: That rational nature is an end in itself means that rational nature, that is, persons, are the ends for the sake of which other things have value and the end for the sake of which rational action is undertaken. FH would then be the formal principle of rational agency because valuing rational nature is a condition of rational choice—roughly, one exercises one’s will by making rational nature one’s end and by valuing it as an end in itself. I find this general approach in the work of Christine Korsgaard and David Velleman. They do not claim that FH is a formal principle in any sense; my contribution is to suggest their approach lends itself to this understanding.

Velleman interprets the idea of persons as ends in themselves as follows:

when Kant referred to persons as ends he was saying merely that they are things for the sake of which other things can have value, as your happiness is valuable for your sake. . . . In his view, persons shed value on other things by making them valuable for the person’s sake. . . . (Velleman 2006: 42, 43)<sup>17</sup>

The statement that a person is an end, I interpret as expressing the fact that we ought to care about some things for the person’s sake, by caring about them out of concern for the person. A person is an end in the sense that he is that for the sake of which—out of concern for which—some things are worth caring about. (Velleman 2008: 191)

<sup>16</sup> But see Herman (2011).

<sup>17</sup> Velleman stresses that that for the sake of which one acts need not be an aim to be produced. In action undertaken for the sake of a person, the person is the object of some attitude or form of concern that motivates the agent to undertake the action. He bases moral constraints of respect for persons on the idea of persons as ends for the sake of which other things have value as follows. That for the sake of which other things have value itself has a value that limits permissible choice, because there is a kind of practical irrationality in subordinating its value (using it, sacrificing it, exchanging its value, etc.) to goals that ultimately only have value for its sake. See Velleman (2006: 43, 88–92; 2008: 192–193).

Korsgaard does not explicitly characterize ends in themselves as ends for the sake of which other things have value, but I believe that she understands the idea of humanity or rational nature as an end in itself in the same basic way. For example:

Kant saw that we take things to be important because they are important to us—and he concluded that we must therefore take ourselves to be important. In this way the value of humanity is implicit in every human choice.<sup>18</sup>

Korsgaard’s view is that valuing humanity (one’s human identity) confers normative force on one’s particular practical identities. Valuing one’s human identity is thus the ultimate source of the value of conforming to one’s particular practical identities, a condition of having reasons for action, and accordingly a condition of (rational) action.

I trust that the idea that persons are ends for the sake of which other things have value is familiar, but let me begin to fill it out through two examples. Consider first a helping action motivated by the judgment that a person’s needs are reasons that make a claim on you. Since the reasons for helping trace back to the value or standing of the other person, you take on the end of helping for the sake of the person and are moved by “respect for the person as an end in himself.” Respect here acknowledges the value or standing of the person as a ground for taking on other concerns (here a concern for the person’s good) and for taking certain facts about the person’s condition as sufficient reasons for action. That is to say that “respect” is a recognition that the value or standing of the person is the basis of compelling reasons to treat that person in certain ways, to give consideration to the person’s interests and good, and so on.<sup>19</sup>

For another kind of example: say I find that a certain area of scholarship interests me in a sustained way, and I take that fact as a sufficient reason (within the obvious moral parameters) to enter the field. For present purposes, I assume that this is a humanly good or worthwhile activity independently of my interest in it and that there are objective standards for how this end is to be pursued. Against this background, my interest is reason enough to devote myself to that field. Having done so, it now matters to me that I master the field and develop a sense of what is important in it, that I immerse myself in significant problems, that I make some original contribution, and so on. It matters to me both that I reach certain levels of achievement and that I live up to certain ideals of scholarship, and I regard my doing so to be objectively worthwhile. What goes into

<sup>18</sup> Korsgaard (1996b: 122).

<sup>19</sup> Here see Velleman, “Love as a Moral Emotion,” in Velleman (2006: 88–93). This case needs to be distinguished from other cases of helping—for example, when one helps to impress another or to incur a debt (for one’s own sake), rather than for the sake of the beneficiary; or when one helps for the sake of the beneficiary, but out of sympathy rather than respect (one is not guided by the idea that the person has a claim on one’s action).



these things mattering to me? Obviously I have reasons to pursue these scholarly ends and to take the values of scholarship seriously. (That these things matter to me may also be reason to hold others to these standards.) Further, how well I do is a reason for certain kinds of affective responses—for example, there are grounds for pride if I make an original contribution, grounds for self-satisfaction if I live up to the ideals of scholarship and for disappointment or self-reproach if I do not, and so on. I am inclined to say that several features of this example (though not all) reflect a value that I place on myself—my initially taking my interest in the field to be a good reason for taking on this end, the way in which these things subsequently matter to me, and my thought that *my* reaching a certain level of achievement is objectively worthwhile. Furthermore, I am inclined to say that if my level of accomplishment matters through a value that I place on myself, then in taking on these ends and acting on the reasons and values of scholarship, I act for the sake of myself as an end.

What interests me in these examples is the implied view about the structure of value and reasons. The familiar thought is that rational action is undertaken for the sake of some person in the sense that the value of some end, or there being reasons for certain actions, are ultimately grounded in the value of some person. The value of rational nature is the terminus of rational support and the formal condition of there being sufficient reason for action. Absent this form of value, there would not be sufficient reasons for action. If the value of persons confers value on specific ends or actions, then in responding to those specific values and reasons, one is acting for the sake of that person, in effect making the person one's end.

Thoughts in this vein permit us to understand FH as the formal principle of rational volition. Rational volition understands itself to specify action through principles that provide sufficient rational support. If the value of rational nature or persons is a formal condition of there being good and sufficient reasons and the terminus of rational support—the end for the sake of which rational action is undertaken—then rational volition understands itself to be for the sake of persons. This self-understanding of rational volition would be expressed by FH. Roughly, you exercise the will by making rational nature or persons your end (in the “end in itself” sense): you find action to have sufficient rational support by reasoning from the value of persons or by framing practical reasoning in terms of that value. A representation of the value of rational nature as an end in itself then figures in all rational volition, and FH tacitly guides all rational choice, though often in defective form. The value of rational nature can be misrepresented (the value of persons can be tied to the wrong capacity or part of the self), it can be represented incompletely (one can value oneself without extending that value to others), agents can reason badly from that value to action, and so on.

The two examples I just gave are intended to fill out a generic version of the idea that persons are the ends of rational action and are not driven by a specific

conception of agency and value. Both Korsgaard and Velleman have conceptions of rational agency that work this idea out in some detail. Let me first sketch some of Korsgaard's views, and then turn to Velleman, who develops a conception quite similar to Korsgaard's. Both offer neo-Kantian conceptions (rather than interpretations of Kant) that suggest renditions of FH as the formal principle of rational volition.

In *The Sources of Normativity*, Korsgaard argues that it is by valuing oneself as a human being that one comes to have reasons for action. That would make valuing one's humanity, or valuing oneself as a human being, a formal condition of rational choice that is implicit in all rational choice.

Korsgaard arrives at this conception of volition through a complex conception of the “reflective structure of human consciousness” that has much built into it (Korsgaard 1996b: 92–3, 103). First, it includes a conception of *negative freedom*: a reflective subject has the ability to step back from any impulse and ask whether it provides a reason (for belief, for action) and it can only move forward (to belief or action) by actually endorsing the impulse (93). This conception of negative freedom leads to a form of *voluntarism*—that a consideration or practical principle provides a reason for a subject only through an act of endorsement. This is not just the weak claim that a consideration or principle can motivate only by, for example, being endorsed or regarded as reason-giving, but rather the more controversial thesis that it gets its normative force, or validity as a reason, from the volitional activity of the agent (e.g., actual endorsement or identification) (121–3, 125, 254).<sup>20</sup> It also leads to a “*positive*” conception of volition according to which the characteristic activity of the will is endorsement and identification with a law or principle (understood as the exercise of the reflective self's authority over the acting self) (104). Volition involves “giving oneself a law,” where the only constraint is that what one wills, or the “law” that one gives oneself, is a general principle.<sup>21</sup> Since to identify with a principle is to regard it as expressive of yourself, the laws that one gives oneself will

<sup>20</sup> In her well-known earlier article “Kant's Formula of Humanity,” Korsgaard ascribes to Kant the view that in choosing rationally we suppose that “rational choice itself makes its object good” and thus is “value-conferring” (1996a: 122). Read in light of some of her later work, her view here, I take it, is that there are features of objects that interest us in them independently of our choices, though they may be features that objects have in relation to our interests, needs, patterns of response, and so on. Choice is an act of endorsement that is necessary for these features to be reason-giving. Further, an individual's choice (assuming that it is consistent with constraints set by the general value of humanity as an end in itself spelled out through application of the Categorical Imperative) makes it, say, a good thing in the judgment of anyone that the individual succeeds in pursuing the end or activity, thus gives others reasons to support her activity.

<sup>21</sup> As she says, “all that it has to be is a law.” See Korsgaard (1996b: 98), and “Morality as Freedom,” in Korsgaard (1996a: 162–7). FUL is the formal principle of volition, conceived as giving oneself a law, because it is the higher order principle of choosing a law, subject only to the constraint that the principle one elects have the form of law. As I understand her view, this constraint is simply that one choose a general principle (and not, for example, the richer constraint that it be a principle that all rational agents can accept as authoritative).



be the basis of a practical identity (or set of practical identities) that give rise to reasons for action and obligations in particular circumstances (101). Given Korsgaard's voluntarism, the normative force of these identities depends on their being endorsed by the agent.<sup>22</sup> "Autonomy is the source of obligation" in that substantive obligations are based on the laws that we (actually) give to ourselves through our willing (104).

Finally, these features of the reflective structure of self-consciousness point to a *conception of human identity*: it is a necessary feature of human agency that we are reflective animals "who need reasons to act and to live" (Korsgaard 1996b: 121), and since reasons depend upon practical identities, we "need to have practical conceptions of our identity in order to act and to live" (121, 129). One's human identity is "a reason" to conform to some of one's practical identifications in the sense that it is a fact about human agency that one must maintain and conform to some practical identity in order to have reasons for action and to exercise one's rational powers. But "it is a reason you have only if you treat your humanity as a practical, normative form of identity, that is if you value yourself as a human being" (121). That is, given her voluntarism, the normative force of this need of human agency depends upon a subject actually treating it as reason-giving. Thus, the normative force of one's particular practical identities, and presumably all reasons for action, depend on actually valuing (identifying with) one's humanity.<sup>23</sup>

If valuing oneself as a human being involves some volitional activity on one's part, it is something that one can fail to do—though at the considerable cost of complete normative skepticism, since then one would not have any reasons for action or see any value in the world. Presumably the default is that agents do value their humanity whenever they act for reasons, even if only implicitly. If ground level reasons depend on giving oneself some law, or

<sup>22</sup> Initially what I am identifying as a conception of negative freedom includes the claim that the reflective subject needs a reason, that is, some general consideration, to go forward. Elsewhere Korsgaard fills this claim out with her argument against "particularistic willing." Negative freedom requires minimally that a subject can move to belief or action only by endorsing some impulse. But a condition of there being a distinction between the acting self who endorses and the impulses within the self is that endorsement or identification be of some general principle that applies to a range of similar cases. For if willing were "particularistic"—if it could consist of endorsing a consideration or an action here and now, with no implications beyond the case at hand—the subject would wholly identify with and in effect be absorbed into the present motive or impulse. But then there would be no distinction between the self and the various desires and impulses in the self, and thus no active self (see Korsgaard 1996b: 225–33; 2009: 72–6.) Thus, the idea that a reflective subject can go forward only through an act of endorsement, when supplemented by the argument against particularistic willing, leads to the need for reasons (general considerations) in order to go on and the idea that volition involves giving oneself a law or general principle.

<sup>23</sup> See for instance Korsgaard (1996b: 125): "Our other practical identities depend for their normativity on the normativity of our human identity—on our own endorsement of our human need to be governed by such identities—and cannot withstand reflective scrutiny without it. We must value ourselves as human."

on some practical identifications, and the normative force of these particular identifications presuppose that one identifies with one's humanity and endorses the general need for reasons, then valuing oneself as a human being is implicit in all rational volition.

A brief digression: what does Korsgaard mean when she talks about valuing oneself as a human being or under one's human identity, and is it true that what appears to be a valuing of *oneself* underlies all volition? A practical identity is a "description under which you value yourself, a description under which you find your life to be worth living and your actions worth undertaking" (Korsgaard 1996b: 101). It is a self-conception that contains a set of inter-related norms, and (I take it) by providing some definition to the self, it makes possible concrete expression of the basic or primitive value that any subject places on him or herself. Valuing oneself involves caring about one's good and thinking that it matters how one's life goes. But in order to have this kind of concern for oneself, one's life needs some shape in terms of which it can go well or badly. Among other things, a practical identity specifies a notion of good and thus provides a substantive way in which to value oneself, giving content to basic self-concern. Valuing oneself as a human being then involves valuing oneself under the description of a reflective animal who needs reasons and stable normative conceptions of her identity. The normative element involves taking oneself to have reason to act and to exercise one's rational powers and thinking that it matters whether one is able to exercise these powers and how one does so. (Roughly it treats exercising one's rational powers as a form of good.) Since it is only an abstract mode of valuing oneself, it leads one to endorse the need for substantive reasons and more particular practical conceptions of one's identity. As one might say, valuing oneself as a human being is the form of taking oneself to have reasons for action and being concerned and thinking that it matters how one's actions and life go.

The thesis that valuing oneself under one's human identity is the ultimate source of value and reasons appears to commit Korsgaard to the view that all rational volition is guided by a basic form of self-concern (that needs specification and concrete expression through particular practical identifications). One might worry that this is false (clearly rational choice can be guided by concern other than for oneself) or that this conception of volition is unduly self-absorbed (should I always be valuing myself or my capacities when I act?). In response, it is important that Korsgaard holds that all reasons are public and are the basis of shared judgments about what is good that hold for anyone (see Korsgaard 1996b: 132–43). Thus, in valuing oneself as a human being one takes it to be good in anyone's judgment that one act and exercise one's rational powers, that one maintain and conform to a stable set of practical identities, and so on. And of course if my endorsement of my human identity is the basis of shared reasons, others' endorsements of their human identity are as well. Proper reflection on one's human identity (aided by some Kantian theorizing) should lead one to

think of oneself as a Citizen in the Kingdom of Ends (1996b: 100).<sup>24</sup> Even so, given Korsgaard's voluntarism, that expression of human identity becomes normative through an agent's endorsement of it, in which case it is still a description under which one *values oneself* as a human being.

The second prong of this worry can be addressed by noting that valuing oneself as a human being remains empty until specified through some particular practical identity and that the identities through which this value can be expressed are entirely open. Valuing oneself under one's human identity seems to include a concern that one live and act well, that one achieve good, or that one act from good reasons, but at a level of abstraction that does not yet determine what counts as good reasons and achieving good. Here it is important that one *can* value oneself as a human being through the identity of a Citizen in the Kingdom of Ends. In that case, one will recognize binding moral obligations as ordinarily understood (by Kantians, at least). One will be committed to having a good will and acting from universally valid principles, one will have the ends of virtue (which include a requirement to have the attitude of respect toward the rational autonomy of others and oneself), and so on. Moreover, one will think that it matters that one act from these principles and values: they determine the good that one thinks it is important to achieve. Thus, the thesis that this formal notion of basic self-concern is a component of rational volition does not imply exclusive or undue influence on oneself.<sup>25</sup>

What then is the functional role of valuing oneself under one's human identity? As we have seen, "autonomy is the source of obligation" for Korsgaard in the sense that what ultimately confers normative force on any set of considerations is endorsement of one's human identity. Furthermore, identification with a set of values and principles gives them a motivational foothold in the subject by making it one's end to act from these values and principles.

There are two general points that I want to draw from Korsgaard. First, her conception of agency makes valuing oneself under one's human identity, or valuing one's humanity, a formal condition of having a reason for action, and thus of the possibility of rational choice. That is because valuing oneself as a human being confers normative force on one's particular practical identities, in relation to which certain ways of living and acting come to have value. In this way the value that one (tacitly) places on one's humanity is the ultimate source of the value and normativity that one finds in the world and the terminus of rational guidance for individual choices. If in valuing oneself as a human being

<sup>24</sup> Cf. also the arguments in "Kant's Formula of Humanity," in Korsgaard (1996a: 119–128).

<sup>25</sup> Concern for animals or for nature presents another obvious problem for the idea that persons or rational nature are the ends for the sake of which rational action is undertaken. The line of thought sketched in this paragraph may provide a way to address this objection, though I won't consider that issue here. (Perhaps Kant was not so far off the mark in classifying duties to animals and to nature under duties to oneself [*MS* 6: 442–3].)

one values humanity generally, then FH is the formal principle of rational volition: it is by (tacitly) valuing humanity as an end in itself that one comes to have reasons and exercises one's will.

Second, if complying with and sustaining some of one's practical identities is a condition of exercising one's rational capacities and of giving concrete expression to the value that one places on one's humanity, or on oneself, then we may say that the particular identities that are fundamental to one's self-conception are adopted for the sake of one's humanity (cf. Korsgaard 1996b: 102–3). In acting from these identities, one acts for the sake of one's humanity. Further, since these identities get their normative force from the value one places on one's humanity (from one's endorsement of one's human identity), the identities adopted for the sake of one's humanity inherit the value that one places on human identity. That suggests that the ends and projects that are central to one's fundamental practical identities have special normative standing, both for oneself and for others. They are sources of unconditional obligations, in Korsgaard's sense—reasons for action that an agent cannot ignore without loss of identity; and they create reasons for others, for example, to respect and to support one's pursuit of such ends.

In recent papers, Velleman develops the view that persons, in virtue of their capacity for autonomy, introduce value into the world—as he says, "persons shed value on other things by making them valuable for the person's sake" (Velleman 2006: 43). Moreover, he regards this fact as a necessary condition of things having value or being worth caring about; "Kant thought that a world without persons would be pitch dark with respect to value" (Velleman 2006: 43).<sup>26</sup> In "Beyond Price" (Velleman 2008), Velleman draws on certain ideas of Harry Frankfurt about caring to lay out the route by which the value of persons introduces value and reasons into the world. Caring is a specific motivational attachment that involves a disposition "to support and sustain [a] desire." Further, caring about certain ends is important for its own sake, independently of the intrinsic worth of what we care about, because "it is the indispensably foundational activity through which we provide continuity and coherence to our volitional lives."<sup>27</sup> Given the needs of human agency, then, we have reason to find ends that we can care about and love.

We need a few assumptions to get persons introducing value: (1) what is distinctive of persons is the capacity for autonomy, where that includes the capacity for self-governance through authoritative reasons, moral constraint, setting goals for oneself, and so on. (2) Persons have value (in virtue of their capacity for autonomy), and the value of persons confers value on their good (where a person's good is "what it makes sense to care about out of an appreciation for"

<sup>26</sup> See also Velleman (2008: 211): "what ultimately makes things worth caring about in the way that entrenches them in a person's good . . . is one's value as a person."

<sup>27</sup> Frankfurt (1999: 160, 162–3); cited in Velleman (2008: 208–9).

the value of the person).<sup>28</sup> (3) A person's good is the realization of his or her autonomy. Now, having continuity and coherence in one's volitional life is a condition of realizing one's autonomy, and having ends that one cares about in sustained fashion is what gives continuity and coherence to one's volitional life. So the value of persons and the good of realizing one's autonomy confer value on any ends that one can care about in a sustained manner and give individuals reasons to find and take on some such ends. At this point in the story, the value of persons introduces reasons or normative considerations: the value of persons makes any ends or activities that *would* unify and give coherence to an individual's volitional life candidates for inclusion in the individual's good and ends in which an individual has reason to take an interest. If the only way for value to enter the scene is through an activity playing this role in individuals' volitional lives, then the value of persons (persons' autonomy) is a necessary condition of value and reasons. (I expect that Velleman would accept some such claim, though I do not know for sure.) Note that specific volitional acts (of endorsement and identification, of choice) have not yet figured in this story, since the ends and activities that have value are those that *could* play a certain role in an individual's volitional life, not those the individuals have endorsed or chosen. In this respect, Velleman's view differs from Korsgaard's, since in her conception reasons and normativity depend on what individuals endorse. But choice comes next in Velleman's story. Once an individual settles on, or even stumbles into, some set of ends that provide continuity and coherence to his volitional life, these ends take on a special status. Those of an individual's ends that play this role in his volitional life are elements of his good.<sup>29</sup> Because they play this role, we may say that such ends and activities are chosen for the sake of the person's autonomy, thus for the person's sake. The value of autonomy now makes the pursuit of these ends objectively worth caring about and gives the person reasons to sustain interest in and take them seriously (cf. Velleman 2008: 211). Presumably this value gives other agents reasons to enable individuals to find and settle on ends that can figure in the realization of autonomy, and

<sup>28</sup> Following Anderson (1993), Velleman holds that something is "valuable if it is worthy of being valued in some way" (Velleman 2008: 200), or if it is the proper object of some evaluative attitude. I take it that he would support the assumption that persons have value through the idea that persons are the proper objects of attitudes such as love and respect, which are responses to persons that lead us to care about other things for the sake of the person. For Velleman the capacity for autonomy that is central to personhood has value in two respects. It is to be valued or respected in oneself, where respect for one's autonomy motivates one to realize that capacity by acting under the guidance of reasons. And it is to be respected and loved in others. The connection is that if there are reasons to respect the autonomy of others, respect for one's own rational autonomy that generates the aspiration to act for reasons motivates one to act on these reasons. See Velleman (2008: 202; 2006: 43–4).

<sup>29</sup> See Velleman (2008: 210): "Not all of [a person's] ends are of significant importance to his good—only those which he cares about in the way that sustains his desire for them. Things are worth caring about in that way because desires so sustained give structure and unity to his life, thereby providing scope for the fullest realization of his autonomy."

once an individual has chosen some such ends, reasons to support his pursuit of these ends. Agents who act on such reasons act for the sake of the person.

We now have the following points from Velleman's view. First, value flows from persons, through their capacity for autonomy and various needs of human rational agency, to ends and activities that can provide continuity and coherence in individuals' volitional lives. These ends and activities are candidates for inclusion in one's good, in which one has reason to take an interest. Second, we may say that ends and activities that do in fact play these unifying roles, through which the individual realizes his good of autonomy, are chosen for the sake of the person's autonomy or rational nature, thus for the person's own sake. These ends are objectively worth caring about because of one's value as a person, and others have reasons to support the pursuit of such ends once chosen.

Finally, as in Korsgaard's view, the value of persons is the formal condition of other values and reasons for action, and the terminus of rational support. For example, it is the formal condition of certain ends and activities comprising an individual's good and for these ends to be worth caring about. The capacity for rational choice is exercised by valuing rational nature or making it one's end (again, in the "end in itself" sense)—by reasoning from the value of rational nature or autonomy to specific ends and actions chosen for the sake of one's rational nature. On this conception of value and agency, FH is the formal principle of rational volition.

How does this general approach to FH fare by the desiderata listed in I.3 above? It satisfies 2. My focus has been on how to understand FH as the formal principle of rational volition, that is, as the principle that constitutively guides or describes rational volition. But the considerations that support this idea establish that rational nature is an end in itself and the substantive value that underwrites moral thought. (If FH is the constitutive norm, it is the regulative norm as well.) Desideratum 2 secures 3a—that FH advances the argument by bringing the moral law closer to feeling and intuition. This approach also fits 4 and 6: if the value of persons is the formal condition of other values and the terminus of rational support, that value would be a necessary ground of there being practical laws. And the end that rational nature sets itself is the "subject of all ends"—that being both the subject that sets ends (ends are set by rational agents) and the subject for whose sake ends are set (ends are set for the sake of persons, rational agents). This end is an "independently existing end," not an end to be produced, since it is the end for the sake of which other ends and activities have value and are chosen; and rational nature *sets itself* this end in the sense that rational volition self-consciously understands the rational nature of persons to be the end for the sake of which other things have value and the end for the sake of which it acts.

The question I have concerns desideratum 1: is FH the *same* formal principle as FUL? Among other things, this worry is fuelled by the way in which desideratum 4 is satisfied. If the value of rational nature or persons is a necessary

condition of practical laws, FH appears to articulate a further condition not contained in FUL on there being principles that provide full rational support or sufficient reasons presupposed by FUL. Here is a way to articulate the worry. Kant argues that a statement of FUL (though not its authority for us) can be derived from the idea of a practical law as a principle that is the basis of fully sufficient reasons for action.<sup>30</sup> I interpret FUL to be the formal principle of free volition in the sense that rational volition understands itself to derive actions from principles that provide sufficient rational support, and that this self-understanding is expressed by FUL. If so, a representation of FUL figures in and tacitly guides all rational volition (e.g., as part of the self-understanding of what volition is, or as a premise in the reconstructed practical reasoning that underlies choice). We have just considered a conception of rational agency in which the value of the rational nature of persons is a formal condition of there being sufficient reasons for action, and rational volition understands itself to be for the sake of rational nature or persons. In this conception of rational agency, a representation of the value of humanity likewise tacitly guides all volition (again, as part of the self-understanding of volition or as a premise in the reconstructed practical reasoning). Are the formal principles of these conceptions of volition the same, or at least different expressions of a single idea? (This is just to raise the standard question of the equivalence of the formulas, but in the context of taking them to be the formal principle of rational volition, according to some conception of rational agency.) The worry is that they are not: that these assumptions about the value of persons do not follow from the idea of practical law from which Kant derives a statement of FUL, and that they add an element to the conception of practical reason and a further condition on the existence of sufficient reasons for action that is not contained in FUL.<sup>31</sup> To dispel this worry one would have to show that the above conception of rational nature as an end in itself (construed as the end for the sake of which other things have value) is implicit in the notion of practical law and sufficient reason, and that I have not done.<sup>32</sup>

<sup>30</sup> See *GMS* 4: 402 and 4: 420–21; and *KpV* 5: 19–30.

<sup>31</sup> This question should not worry Korsgaard or Velleman since they do not think that FUL and FH are equivalent. See Korsgaard (1996a: 143–4, 151–4) and Velleman (2006: 40).

<sup>32</sup> Kant thought both that a statement of FUL can be derived from the very idea of a practical law and that this principle has substantive implications for choice and action. To address this issue about equivalence, it is not enough to show that the existence of practical laws and sufficient reasons presupposes that rational nature is an end in itself: one needs to show that this conception of rational nature as an end in itself (construed as the end for the sake of which other things have value) is derivable from the very idea of a practical law or sufficient reason. It may be that there would not be sufficient reasons for choice unless rational nature were an end in itself, but that the absolute value of rational nature is not derivable from or contained in the idea of sufficient reason; that an independent argument is needed to show that rational nature is an end in itself. (And in fact it is natural to read the *Groundwork* as introducing new material at *GMS* 4: 428ff. that goes beyond what has been introduced so far, that is a condition of there being practical laws

### III: A Second Formal Reading: The Formal End of Practical Reasoning

According to the second approach, that rational nature is an end in itself means that the formal end of practical reason is its own proper exercise, as defined by FUL. FH is then a formal principle because it is strictly equivalent to FUL. This reading is proposed by Stephen Engstrom in his recent book *The Form of Practical Knowledge*.<sup>33</sup> (It has also been suggested in passing by Barbara Herman.<sup>34</sup>) To explain Engstrom's approach, I need to give some background on his interpretation of the Categorical Imperative.

Engstrom connects the Categorical Imperative with reason by showing how it can be unfolded out of the idea of "practical knowledge" of what is intrinsically good, where the idea of knowledge introduces the conditions of universal validity expressed in FUL. He understands practical reason or the will as the capacity for practical knowledge of the good—roughly, knowledge of what ought to be that has the capacity to make its object real. For Engstrom, all rational volition is based on a judgment about what is good, where the formal feature of judgment (theoretical or practical) is that it understands itself to be making a universally valid claim, a claim to knowledge. Engstrom distinguishes two notions of universal validity. First, a judgment understands itself to be "subjectively universally valid"—to hold *for* all judging subjects. Second, it understands itself to be "objectively universally valid," or to hold *of* all objects that fall under the concept employed in the judgment. The Categorical Imperative is based on the formal features of specifically practical knowledge. In practical judgment, the subject and object of the judgment are identical: "practical knowledge is always knowledge cognizing subjects have of what *they themselves* are to do" (Engstrom 2009: 121). Thus in practical judgment subjective and objective universal validity coincide: they are judgments about what any subject in the relevant conditions is to do that are valid for all subjects with the capacity for practical knowledge. This "double universal validity" is the "form of practical knowledge"—the condition that practical judgments must satisfy to count as genuine knowledge of what is good (122–4). Accordingly

and sufficient reasons, etc.) Elsewhere I've argued that the condition on choice imposed by FUL is that a principle can be willed as universal law for agents with autonomy, where a conception of autonomous agency is a source of moral content. FUL may be understood in this way because some notion of autonomous agency is implicit in the idea of a practical law. Specifically it is part of the idea of a practical law that the agents subject to such principles must be regarded as their legislators (*GMS* 4: 431)—that is, they must have the practical and legislative capacities that go into Kant's conception of autonomy. If practical laws govern the conduct of agents with autonomy, then they must be universal laws for agents with autonomy (see "Agency and Universal Law" in Reath [2006, especially 204–8 and 211–20]). Perhaps one can use this argument to derive the value of rational nature as an end in itself from the idea of practical law or sufficient reason, say, as part of the self-conception of any agent with the capacity to act from a practical law. But space does not permit pursuing that line here.

<sup>33</sup> Engstrom (2009).

<sup>34</sup> Herman (2007: 250–53).

Engstrom interprets FUL as the imperative: to act from maxims such that all practical subjects can agree that any subject in the same conditions is to act from the principle (124–5, 221).

Two more preliminary points: First, Engstrom argues that, since the formulas of the CI are equivalent, this “double universality” is implicit in all of them. The different formulas all express the form of practical knowledge, though they highlight different aspects. FUL, which Kant connects with the concept of a practical requirement and subjection to duty, stresses the objective universal validity of a practical law—that a practical law is to govern the conduct of all agents who fall under it. The focus on humanity or rational nature articulates the idea that moral agents are “cognizing subjects” with the ability to make and act on the basis of judgments about good; thus it highlights subjective universality, that all judging subjects must be able to agree with and hold a valid practical judgment. The Formula of Autonomy (FA) highlights the coincidence of these two forms of universal validity—that “the subjects to whom law is given are necessarily the subjects in and through whom it is given by the practical reason . . . that they share in common” (Engstrom 2009: 136; cf. 150),<sup>35</sup> Second, Engstrom accepts a strong guise of the good thesis: all practical judgment, that is, all rational volition, understands itself to satisfy this double universality and carries what he calls the “presupposition of universality.” It understands itself to be directed at objects taken to be good in this strong sense. That makes FUL the constitutive guiding principle of rational volition in the sense I have discussed.

Turning now to Engstrom’s interpretation of FH, he takes “rational nature” or “humanity” to be the capacity for practical knowledge (i.e., practical reason), which is exercised through FUL (2009: 167–8). To treat rational nature as an end is to represent that capacity as an end in all practical judgment (170–71). Two points are involved here, that all practical judgments contain a *representation* of rational nature and that it is represented as an *end*. First, it is part of the form of practical knowledge that a practical judgment is about what some practical knower (i.e., an agent with rational capacity) is to do. For example, in judging that an action is choiceworthy or good, I make a judgment about what I, as a rational agent, am to do. The practical judgment contains a representation of rational nature because it represents oneself (or the agent about whom one makes the judgment) as having the capacity for practical knowledge—the capacity to determine oneself to act from a judgment about good that satisfies

<sup>35</sup> For Engstrom, the idea that practical laws are “self-legislated” appears to involve two elements. First, the agents subject to practical laws have the shared capacity for practical knowledge, the capacity to make judgments about objective good, through whose exercise the content of moral law is determined. Second (and perhaps more importantly), since these judgments are *practical*—that is, they have the capacity to bring their objects about—it is through the judgments of such subjects that practical laws are efficacious in determining conduct, and are thus real laws. As Engstrom says: “For a law whose efficacy, and so whose very being as a law, depends on its being known by those whose existence it can determine is precisely a self-legislated law” (2009: 136).

the presuppositions of universality. Here Engstrom says: “Humanity is thus represented in the subject position in all practical knowledge, prior to all acts of practical predication through which particular ends are adopted” (2009: 171).

Second, the capacity for practical knowledge is represented as its own end—again as part of the form of practical knowledge. A practical judgment understands itself to be making a universally valid claim and it sustains itself through self-consciousness of its own validity. It aims at and takes itself to be in agreement with all other practical judgments (including those of other judging subjects), it draws support from other practical judgments, and it withdraws its claim if these conditions are not met. Engstrom writes:

humanity is always represented in such cognition as already actual and self-sustaining in and through such cognition itself. As Kant says, it is conceived as “self-standing” [*GMS* 4: 437]. Hence in practical cognition humanity is represented, not as to be produced, but as to be sustained, both in a negative sense (as not to be hindered) and also positively (as to be furthered) so far as practical cognition . . . is capable of developing and perfecting itself. (2009: 172)

His point, I take it, is that practical judgment is guided by its formal or constitutive aim of satisfying the presupposition of universality, and moreover it understands itself to be always on the way to satisfying this condition. Practical knowledge is not some distant aim of judgment: a practical judgment takes itself to be an instance of practical knowledge. In that sense, humanity or rational nature is not an end to be produced, but a capacity to be “sustained.” Engstrom’s larger claim, if I understand it, is that rational nature (practical reason) is represented as an end in practical judgment and knowledge in the sense that such cognition understands itself to have, thus is self-consciously guided by, the formal end of satisfying its own internal norm (the conditions of universality). That is, the claim that rational nature is an end in itself is—at least initially—the claim that rational nature has the formal end of its own proper exercise, and thus is its own end. Since all exercises of rational nature are tacitly guided by its own formal principle, this is an end that it is always on the way to actualizing; in that sense this end is “self-standing.”<sup>36</sup>

<sup>36</sup> Engstrom relies on a particular understanding of what ends are: an end is “represented in practical knowledge as being for its own sake, . . . [F]or something to be deemed good as an end, or for its own sake, is for it to be represented in a practical judgment as furthering itself . . . [A]n end always sustains itself” (2009: 74–5). As I understand his view, ends sustain themselves through rational agents’ representations of them (through rational agency). An end has features that, when represented by a rational subject, lead the subject to maintain the thought of the end. That is to say that the subject takes pleasure in the thought of the end, and the representation of the end leads to active interest in its actuality. What matters for our purposes is that practical judgment takes itself to satisfy the presupposition of universality and sustains itself through its awareness of satisfying these presuppositions. Since it is represented in practical knowledge as sustaining or furthering itself, it satisfies the concept of an end, and indeed is its own end.

The capacity for practical knowledge is properly exercised when a judgment in fact has the double universality that is the form of practical knowledge. The judgment will then “agree with humanity” because it fully conforms to the principle that defines humanity, or the capacity for practical knowledge. That means that one conforms to FH by conforming to FUL (Engstrom 2009: 173).

One virtue of Engstrom’s interpretation is that it makes FUL and FH strictly equivalent. FH adds no fundamentally new ideas, since it is just another expression of the form of practical knowledge. It simply articulates a notion implicit in FUL: that the agents subject to moral requirements are “practical knowers” with the capacity to make judgments about good, and that it is a condition on any valid practical judgment that all such subjects can agree with or hold that judgment.

An additional virtue is that this interpretation aligns the idea of rational nature as an end in itself with the absolute value of the good will in an interesting way.<sup>37</sup> Bearing in mind that practical reasoning is the basis of volition, the formal end of satisfying the conditions of universality would appear to be that of having a good will. So the idea that rational nature is its own end, as understood by Engstrom, amounts to the idea that it has the formal end of good willing. Further, the formal aim of good willing has unconditional authority over all exercises of the will since having that aim is a condition of willing. If you don’t aim to satisfy the internal norm of volition (the conditions of universality), you are not exercising the will. As we might say, no other aim can be put in the place of this formal end, and it is never to be abandoned for any other end. Thus, taking the formal end of rational volition to be good willing seems close to building recognition of the absolute value of the good will into the self-understanding of rational volition. Here consider the “practical” reading of the absolute value of the good will suggested by Thomas E. Hill, Jr.<sup>38</sup> To hold that the good will is good unconditionally and without qualification is to say that it is always worth choosing or maintaining in all circumstances, and never to be abandoned for any other kind of good. And one chooses or maintains a good will by recognizing the priority of moral reasons and conformity to universal law. The point that I want to make here is that assigning rational volition the formal end of good willing gives that end authority over all exercises of the will, and that amounts to building a thin version of the recognition of the absolute value of the good will into the form of volition.

A third strength is that this approach leads to a satisfying reading of Kant’s claims that an end of absolute value is a necessary condition of a categorical imperative (desideratum 4). Given the fact that Kant treats FUL as a principle that can stand on its own, why is an end of absolute value a *necessary* ground of a categorical imperative? But what if rational nature were not an end in itself in the sense we are considering—that is, what if rational volition did not have

the formal end of good willing, or that this formal end did not have authority over all exercises of the will and that some other aim could be put in its place (that is, it did not have “absolute worth”)? Then there would be no authoritative reason to conform to the conditions of universality and they would not be the basis of true practical laws. (Rational volition would be free, as it were, to aim at something other than conforming to the conditions of universality, in which case it would have no authoritative standard and all practical principles would be conditional.) The “absolute worth” of rational nature is the ground of practical laws because it amounts to the recognition of the authority of the internal principle or formal end of rational volition.

But Engstrom’s interpretation does have one feature worth noting. The requirement to treat rational nature as an end amounts to the injunction to agree with its formal end by reasoning in ways that do in fact satisfy the presupposition of universality. This is reflected in a passage cited earlier, where Engstrom writes that humanity (the capacity for practical reason) is “to be sustained, both in a negative sense (as not to be hindered) and also as positively (as to be furthered)” (2009: 172). What is to be sustained, I take it, is the proper use of practical reason. Since rational agents are already on the way to reasoning properly, (negatively) one should avoid influences that impede its proper use and lead to judgments that conflict with those of others; and one should perfect one’s rational capacities, so that one’s reasoning displays “positive agreement with humanity as an end in itself” and agrees with the practical judgments of others. (*GMS* 4: 430; II. 54).<sup>39</sup>

As Engstrom interprets it, the value of rational nature as an end in itself is not on its face the rich notion of respect for persons in the standard intuitive sense that underlies much ordinary moral thought. The rich notion requires giving certain forms of consideration to persons (including ourselves) on the receiving end of our actions and treating them in certain ways. But here treating rational nature as an end in itself requires a certain attitude toward that capacity: one is to recognize the authority of its formal end (exercising the capacity according to its own internal norm).<sup>40</sup> Note that the negative and positive

<sup>39</sup> I gather that “sustaining humanity” involves judging in ways that are consistent with the aim of shared practical judgments. Negatively one is to avoid practical judgments that lead to contradictions in conception; positively one’s judgments should agree with judgments about the goodness of self-perfection and own happiness to which all agents are committed, thus avoiding contradictions in will. I’m grateful to Steve Engstrom for comments here.

<sup>40</sup> What has absolute worth or is to be “respected” as an end in itself is a capacity that one possesses—the capacity for practical reason, understood through its formal principle. There is, of course, precedent for this reading in the texts. At *GMS* 4: 428 and in the four examples (*GMS* 4: 429–31), it is clear that persons and their rational capacities are ends in themselves. But in a key discussion of “respect,” Kant notes that “the object of respect is therefore simply the law” and that honorific respect for a person’s accomplishments is “respect for the law . . . of which he gives us an example” (*GMS* 4: 401n). For an interesting discussion that understands respect for the moral law as respect for the idealized rational will that is the essence of the person and a law for the empirical self, see Velleman (2006: 77–81).

<sup>37</sup> Engstrom does not make this point, and I go beyond his account here.

<sup>38</sup> See “Is a Good Will Overrated?” in Hill (2002: 37–60, especially 42–3, 52).



aspects of respecting the formal end of practical reason to which Engstrom points are a step removed from the negative and positive duties that Kant uses to illustrate the FH. The focus of these examples—negative duties proscribing suicide, deception, interference with the rights of others, and the positive duties of virtue (self-perfection and mutual aid)—is the proper attitude toward and treatment of persons and their rational capacities.

Is this a problem? Perhaps it is not, because the proper use of practical reason leads to choice that respects persons in the standard intuitive sense through the condition of subjective universal validity—that all rational agents be able to agree with or hold a valid practical judgment. (My judgment that I may deceive for self-interest cannot be generally shared, in particular by the person I deceive.) That subjective universal validity is a formal feature of rational volition that limits its proper exercise translates into the idea that persons as rational agents are “supreme limiting conditions of the freedom of action of very human being” and of “all subjective ends” (*GMS* 4: 431).

Regarding the six desiderata, Engstrom’s approach does well by 6. The end that rational nature sets itself is its own proper exercise. It “sets itself” this end in that it is constitutively guided by its self-understanding of its formal end, and this is a “self-standing” end that it is always on the way to actualizing. As just discussed, it provides a satisfying, if spare, treatment of 4. It does very well on 1, since it treats FUL and FH as strictly equivalent. In particular, FH is as much a statement of the formal principle of morality and the formal principle of rational volition as is FUL. My question concerns desiderata 2 and 3a: is the link between Engstrom’s interpretation of rational nature as an end in itself and the substantive value of persons sufficiently direct? Connecting the moral law with the formal end of good willing does not exactly bring it closer to intuition and feeling.

#### IV. Conclusion

I hope to have made the case that FH can be construed as a formal principle. The question then is how best to do this. If the approach drawn from Korsgaard and Velleman is to preserve the equivalence of FUL and FH, one needs to show that the key ideas of FH, for instance that persons are sources of value and that rational choice is for the sake of persons, are implicit in FUL. Whether that can be done is a matter for another occasion.

Since Engstrom’s interpretation of rational nature as an end in itself is not on its face the rich notion of respect for persons, the question is whether it preserves the substantive value that underwrites moral thought in the right way (desiderata 2 and 3a). Now this may seem like a quibble, since the formal end of rational nature, once fleshed out, certainly gets us to the value of persons. The conclusion that I think we should draw is that this approach does satisfy 2 and 3a, and in a

way that proves instructive. The value of rational nature as an end in itself (the value of the formal end of conforming to its internal norm) is as abstract as the requirement of conformity to universal law—unsurprisingly, since it is the same idea slightly redescribed. But one aspect of this end is the condition of subjective universal validity—that all subjects be able to hold a valid practical judgment. This idea gives some precision to the idea of respect for persons that underlies ordinary moral thought—for example, setting out an ideal of justifiability as what is owed to persons specifically as rational agents with autonomy. Furthermore, FUL leads to a set of duties and principles whose content is to respect the rational nature of persons. This indicates that FUL, at different levels of generality, contains the substantive value that is clearly worthy of our allegiance. Furthermore, since the value of respect for persons that drives much ordinary moral thought turns out to be implicit in FUL, Kant’s foundational project would show that value to have a genuine basis in reason. In sum, what we might call “the formal formal approach” does surprisingly well on all of the desiderata.

#### Bibliography

- Anderson, E. (1993) *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.
- Engstrom, S. (2009) *The Form of Practical Knowledge: A Study of the Categorical Imperative*. Cambridge, MA: Harvard University Press.
- Frankfurt, H. (1999) *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.
- Herman, B. (2007) *Moral Literacy*. Cambridge, MA: Harvard University Press.
- Herman, B. (2011) “The Difference that Ends Make,” in J. Wuerth (ed.), *Perfecting Virtue: New Essays on Kantian Ethics and Virtue Ethics*. Cambridge: Cambridge University Press.
- Hill, T. E., Jr. (1992) *Dignity and Practical Reason in Kant’s Moral Theory*. Ithaca, NY: Cornell University Press.
- Hill, T. E., Jr. (2002) *Human Welfare and Moral Worth: Kantian Perspectives*. Oxford: Oxford University Press.
- Kant, I. (1996) *Practical Philosophy*. Tr. and ed. M. J. Gregor. Cambridge: Cambridge University Press.
- Korsgaard, C. M. (1996a) *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- Korsgaard, C. M. (1996b) *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Korsgaard, C. M. (2008) *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.
- Korsgaard, C. M. (2009) *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- O’Neill, O. (1989) *Constructions of Reason: Explorations of Kant’s Practical Philosophy*. Cambridge: Cambridge University Press.



- Rawls, J. (2000) *Lectures on the History of Moral Philosophy*. Ed. B. Herman. Cambridge, MA: Harvard University Press.
- Reath, A. (2006) *Agency and Autonomy in Kant's Moral Theory*. Oxford: Oxford University Press.
- Reath, A. (2010) "Formal Principles and the Form of a Law," in A. Reath and J. Timmermann (eds.), *A Critical Guide to Kant's "Critique of Practical Reason"*, 31–54. Cambridge: Cambridge University Press.
- Scanlon, T. M. (1998) *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Velleman, J. D. (2006) *Self to Self*. Cambridge: Cambridge University Press.
- Velleman, J. D. (2008) "Beyond Price," *Ethics* 118(2): 191–212.

{ 10 }

## Kant's Grounding Project in *The Doctrine of Virtue*

Houston Smit and Mark Timmons<sup>1</sup>

Kant's project in *The Metaphysics of Morals* is to set forth and defend a wide-ranging system of general duties by deriving them from a single moral principle, the categorical imperative. In the introduction to this work, Kant claims that the moral law expressed by the categorical imperative "affirms what obligation is" and then he remarks that "the simplicity of this law in comparison with the great and various consequences that can be drawn from it must seem astonishing at first..." (*MS* 6: 225).<sup>2</sup> The great and various consequences in question compose the system of duties that Kant divides into juridical duties and ethical duties according to the type of lawgiving associated with the duty, and which are treated respectively in Parts 1 and 2 of the *The Metaphysics of Morals*: the *Rechtslehre*, or *Doctrine of Right*, and the *Tugendlehre*, or *Doctrine of Virtue*.

Our focus in this chapter is on the relationship between the categorical imperative and the "great and various consequences" featured in the *Doctrine of Virtue* (*DV* henceforth) that Kant attempts to derive from this imperative. (Call them "Kant's derivations.") These derivations feature the Formula of Humanity (FH) of the categorical imperative, which commands individuals to treat humanity as an end in itself, never merely as a means.<sup>3</sup> We construe

<sup>1</sup> This article is thoroughly collaborative. Order of authorship is alphabetical. We would like to thank Sorin Baiasu, Cole Mitchell, and especially Oliver Sensen and Robert Audi for helpful comments.

<sup>2</sup> Passages from Kant's writings in English are from the various translations in the Cambridge edition of the works of Kant listed among the references (Gregor 1997; Heath and Schneewind 1997; and Zöllner and Loudon 2007). In this particular passage, Kant is discussing the universal law formulation of the categorical imperative. Given the alleged equivalence of the various formulations and the fact that in the *Doctrine of Virtue* Kant almost exclusively employs the humanity formulation in his various derivations, his claim about the simplicity of the formula and the many consequences that can be drawn from it presumably applies also to the humanity formulation.

<sup>3</sup> The lone notable exception is Kant's argument for the duty of beneficence in which he employs the universal law formulation. We have more to say about this matter below in section IV. For a nuanced treatment of the notion of treating someone merely as a means, see Audi (n.d.).