Canadian Journal of Philosophy (2025), 55: 7-8, 560–576
doi:10.1017/can.2026.10090

SYMPOSIUM ARTICLE

The Will and the Good

T. A. Pendlebury

University of Chicago, Chicago, USA
Email: tapendlebury@uchicago.edu

Abstract
Recent decades have witnessed sharp disagreement about the character of Kant's metaethical position. Constructivists have read him as thinking of moral value, against the example of his rationalist predecessors, as constructed by acts of the human mind. Realists have read him as thinking of such value, in the company of these predecessors, as constituted prior to human nature altogether. I argue that both camps have mischaracterized Kant's position, in particular as it contrasts with that of the rationalists, and that the correct characterization depends on an underappreciated distinction between two forms of constitutivism. While the rationalists follow a medieval precedent in thinking that the human will is by nature the power to love and pursue a good constituted prior to it, so that they are aptly described as endorsing a heteronomist constitutivist conception of the relation of the will to the good, Kant defines the good in terms of the will as the object to which its internal standard directs it, thereby embracing an autonomist constitutivism.

Keywords: the will; the good; constructivism; realism; Kant's metaethics

1. Introduction

It was John Rawls who first applied the label "constructivism" to Kant's moral philosophy. Since then, a number of commentators, including such figures as Christine Korsgaard, Onora O'Neill, and Andrews Reath, have followed suit, though sometimes they have preferred the terms "procedural realism" and, more recently, "constitutivism."[1] Commentators otherwise inclined, like Karl Ameriks and Allen Wood, have argued, to the contrary, that Kant is not a constructivist or merely procedural realist but, in contemporary terms, a substantive value realist, in allegiance with his rationalist predecessors against the sentimentalism of the early modern empiricists.[2]

As this last remark indicates, this debate has to do both with the classification of Kant's position according to the taxonomy of contemporary metaethics and with the extent to which he aims to make a break with rationalist ethics and moral psychology. Since ethical rationalism is often associated with contemporary realism and sentimentalism with non-realist positions of various kinds—like expressivism—it is natural that these two questions might be treated at once. But they are nonetheless distinct, and their relation is a complex matter. In this essay I set the contemporary question aside—so that my use of the words "realist" and "constructivist" will be undergirded by no developed understanding of that distinction—and focus on the historical one.

[1]See, e.g., Rawls (1980, 2000), Korsgaard (1996a, 2009), Korsgaard (1996b), O'Neill (2012), and Reath (2006). For a more recent approach to the debate, see Schafer (2015).

[2]See, e.g., Ameriks (2003), Wood (2008), and Kain (2017).

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Canadian Journal of Philosophy, Inc. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

On this question, I think that no party to this debate has managed to capture the heart of the matter which it concerns. My sympathies lie with the Rawlsians to the extent that they say, against realist readers, that Kant is just as concerned to reject rationalism as to reject sentimentalism. But their various descriptions of this rejection have proved vulnerable to realist rejoinders, with much of whose substance I am also sympathetic.

These various descriptions might be arranged according to the three formulations of the Categorical Imperative: the Formula of Universal Law (or "FUL"), the Formula of Humanity (or "FH"), and the Formula of Autonomy (or "FA"). Rawlsian interpreters have claimed to find grounds for their reading in connection with all three. I, on the other hand, think that if we are to find a radical break with rationalism, there's only one place to find it: in connection with the FA and, in particular, in Kant's contradistinctive discussion of systems of heteronomy and his own system of autonomy. In particular, I shall argue that Kant's embrace of autonomy makes him a kind of constitutivist, though not a proceduralist or a constructivist. Nonetheless, just as realist readers tend not to appreciate the importance of Kant's constitutivism, Rawlsians tend not to appreciate that constitutivism is not, in the modern period, peculiar to Kant. It is endorsed by his German rationalist predecessors, who in this follow the example of most medieval Aristotelians. Thus, I shall argue, the distinction between heteronomy and autonomy that Kant has chiefly in mind is not a distinction between realism and constitutivism but a distinction between two kinds of constitutivism: a pre-Kantian kind, which says that our function is to serve a normative order outside us, and Kant's Copernican kind, which says that our rational activity is to be explained by nothing external to our own volitional nature.

I shall begin with some brief negative remarks about the FH (Section 2), against the background of which I shall proceed to the FA, where, as I have said, I think the action is, and where the bulk of my argumentative work will be localized (Sections 3–6). One topic I shall not address in any detail is that of the kind of constructivism associated with the FUL, which lies in Rawls's notion of a "Categorical Imperative (CI)-Procedure." This may seem outrageous, inasmuch as this is, historically speaking, the first avatar of the constructivist reading. But the trouble is that it is not a very good one. The constructivist target of realist criticisms of his Kant interpretation is the position that the CI-procedure or its enactment is the explanatory ground of the permissibility of the maxims that it certifies. But Rawls often describes the procedure as a procedural representation of something else, "the requirements of practical reason" (Rawls, 2000: pp. 163–164). Thus, "the judgments that result from a correct use of the procedure are correct…*because* they meet all the requirements of practical reason" (Rawls, 2000: p. 239; emphasis mine). Moreover, he describes "rational intuitionism," which he contrasts with Kant's position, as the view that moral facts or properties are independent of practical reason itself, not just given independently of a procedure or its enactment (Rawls, 2000: p. 241). Thus, I do not think it is clear that Rawls's procedural constructivism, whatever it is, includes the assignment of any explanatory role to the CI procedure as such, as opposed to practical reason, in connection with the rightness or wrongness of actions. Since, however, this is a topic in the interpretation of Rawls that deserves its own sustained treatment, I set it aside.

2. The FH

But what would it be, in any case, for practical reasons to enjoy explanatory pride of place in a theory of the rightness and wrongness of actions? There are, *inter alia*, both realist and constructivist answers to these questions, which might be formulated in terms of corresponding treatments of Kant's FH, to which I now turn.[3] Kant's remarks in the transition from the FUL to the FH in the

[3]Kleingeld and Willaschek (2019) argue that there is an answer that is neither realist, nor constructivist, nor even constitutivist.

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

Groundwork of the Metaphysics of Morals have been seized upon by realist and constructivist readers alike. Here he claims that if "there were something the *existence of which in itself* has an absolute worth," then "in it, and in it alone, would lie the ground of a possible categorical imperative." He then claims that there is such a thing: "every rational being *exists* as an end in itself." He says that, unlike "objects of the inclinations," and "[b]eings the existence of which rests… on nature," which "have only a relative worth," rational beings are "*objective ends*," without which "nothing of *absolute worth* would be found anywhere," so that "no supreme practical principle for reason could be found anywhere" [*Grundlegung* 4:428].

Allen Wood, who will serve as my paradigm FH-realist, is moved by Kant's claim to the effect that if there is a categorical imperative, its ground is an end in itself. He takes this to mean that the categorical imperative, as a principle of the rational will, is grounded in a kind of value that some beings—beings with practical reason—just have. He takes Kant to argue that this status is a condition on the value of any of our ends: all conditional worth must have its condition, so that if you go far back enough, you must reach something with absolute worth (Wood, 2008: p. 92). So as long as you accept that your ends have conditional value, if Kant's argument works, you must accept that you and all other rational beings have the status of ends in themselves.

Christine Korsgaard, who will serve as my paradigm FH-constructivist, agrees that rational beings have a status that conditions the value of anything else and that this status is the ground of the categorical imperative. But Korsgaard does not think that anything *just is* valuable. Rather, we value things and in so doing confer value on them. And insofar as we confer value on anything, we confer, or are committed to conferring, the status of end-in-itself on ourselves: "[y]ou must value your own humanity if you are to value anything at all" (1996: p. 123).[4] Both Wood and Korsgaard thereby endorse regressive readings of the transitional remarks, as expressing an argument that concludes that something—rational beings—has a status which is, in turn, the explanatory ground of the moral law. But they understand this status in different ways: in a value-realist way, in Wood's case, and in a value-constructivist way, in Korsgaard's. The difference comes to this, that Korsgaard takes our value to depend upon some act that we perform or are committed to performing, whereas Wood denies that there is any such dependence (Wood, 2008: p. 92).

But it is not obvious that Kant holds the value of humanity to be the explanatory ground of the law in either of these senses. Consider the following alternative reading of the transitional remarks. We know that the law says to act only on that maxim which you can at the same time will that it should become a universal law. But what *is it* so to act? How can we describe this kind or mode of action in such a way as to make it both motivationally and normatively intelligible that anyone would ever so act? Kant's answer is that to act only on that maxim which you can at the same time will that it should become a universal law is to treat all humanity never merely as a means but always as an end. To say this is to bring the law "closer to intuition" and gives us "access" to it [*Grundlegung* 4:437]. Thus, on this proposal, humanity's status as an end in itself is not an explanatory ground of the law—a fact that explains why we are bound by it—but rather the meaning and spirit of the law and thereby elucidates what we are doing when we act from it.

But suppose we say that, in preference to the positions of Wood and Korsgaard. Then what are we to say about the explanatory ground of the law in practical reason? If it is not the value (or valuing) of practically rational beings, what could it be? It could be practical reason, or the rational will, itself: practical reason itself, rather than any value it might be said to have. My own preference is for a *constitutivist* specification of this idea on which the law articulates an *internal standard* of the will's exercise, so that to the extent that you act as the law prescribes, you exercise

[4]This does not appear to be her position at the time of her earliest treatments of this argument, but it is her position at least as early as (Korsgaard 1996a: p. 407). It is elaborated in *The Sources of Normativity*, in which she also attributes it to Kant (Korsgaard 1996b: p. 122).

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

your will well, and to the extent that you do not, you exercise it badly.[5] When Kant says that an imperative is "possible" [*Grundlegung* 4:417], I understand him to claim that it expresses an internal standard for volitional activity. And that is why it is binding on volitional activity. Evidence for this is to be found in the character of the arguments he furnishes for his possibility claims, both in the case of the hypothetical imperative—if you do not take sufficient means to your ends, you fail on your own terms, because to set an end *is* to will to make yourself the cause of something, and taking sufficient means is nothing other than making yourself the cause [*Grundlegung* 4:417]—and in the case of the categorical imperative—if the maxims you will are not universalizable, you fail on your own terms, because to will a maxim *is* to institute a law for the rational will, but only if your maxim is universalizable can it be a law [*Grundlegung* 4:446–7].

Now, I have just opposed the constitutivist conception of the groundedness of the law, claiming it as my own preference, to the value-realist and value-constructivist views of Wood and Korsgaard. But, funnily enough, this constitutivist story is one which Wood and Korsgaard themselves explicitly affirm, namely, that the law is grounded in the nature of the will, and in precisely the sense I mean. Both of them explicitly endorse a model of the relation of the law to the will in, as Wood puts it, the terms of "the natural teleology involved in…proper functioning" (Wood, 2008: p. 115); see, for a similar characterization, Korsgaard (2009: p. 59ff.). But if I have billed constitutivism as an alternative to value realism and value constructivism, how could each of them endorse both it and one of these other positions?

As far as I can tell, neither Wood nor Korsgaard gives a clear answer to this question. The most straightforward answer on Wood's behalf, it seems to me, is that it belongs to the will's function to observe a value independent of this function, not in the sense that there would be anything with this value were there nothing with this function—since it is only things with this function that are said to have the value in question—but in the sense that what it is for them to have this value is not exhausted by the fact that it is the function of the will to treat them in a certain way (contrary to the proposal of, e.g., Sensen (2018)). In Korsgaard's case, there are two alternatives that her texts seem, at first glance, to make available. On the first, her value constructivism is actually not explanatorily additional to her constitutivism after all. Rather, she holds that part of what it is to exercise the will well is to value humanity in yourself and everyone else. On the second, like Wood, she holds that it belongs to the function of the will to observe a kind of value independent of this function in the sense that what it is to have this value is not exhausted by the fact that it is the function of the will to treat beings with it in a certain way. But she has a constructivist, as opposed to a realist, conception of this value, in the sense that she thinks it is conferred or constituted by acts of valuing or by our commitment to performing such acts. These two Korsgaards thus endorse different orders of explanation. The first Korsgaard says that you are committed to valuing humanity because the internal standard of the will tells you to value it. The second Korsgaard says that the internal standard of the will tells you to value humanity because you are committed to valuing it.[6]

I draw out these speculative interpretive possibilities in service of the following point. It is easy to see why Wood, given his value-realist explanation, might insist, against the second Korsgaard's

[5]See Fix (2021) for a clear statement of internal-standard capacity-constitutivism. See Schafer (2019, 2021) for the application of capacities-constitutivism to Kant. See Pendlebury (2025) for a capacities-constitutivist reading of the relation of morality to happiness in Kant's moral psychology and Fix and Pendlebury (forthcoming) for a capacities-constitutivist reading of the relation of Kant's notions of practical reason, the will, and choice.

[6]For this second Korsgaard to be distinct from the first Korsgaard, it must be that she holds that the sense in which you are committed to valuing your humanity is not exhausted by its being the case that valuing your humanity is part of what it is for you to exercise your will well. For the second Korsgaard to be a constructivist and not a realist, it must be that she holds that the reason you are committed to valuing your humanity is not that it is valuable. (For example, she might argue that it is irrational to value anything unless you also value your capacity to value it.) It seems to me that the second Korsgaard position is philosophically stable, but it might be argued that either of these two further distinguishing commitments is untenable. Whether that is the case is beyond my remit here.

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

value-constructivist explanation, that Kant is a realist and not a constructivist—this notwithstanding the constitutivism that they both endorse. But, surprisingly, it is not merely on the basis of the positions on which they disagree that they come to different conclusions on the realism-versus-constructivism question but also on the basis of their common constitutivism. That is, from the claim that both Korsgaard, Wood, and I all accept—that the moral law is grounded in the nature of the will in the sense of expressing its internal standard—and from that claim alone, Korsgaard and Wood draw opposite conclusions concerning Kant's metaethics. While Wood thinks that Kant's depiction of the law as grounded in the nature of the will makes him a realist, Korsgaard thinks that it makes him something else.

How could Korsgaard and Wood reach the opposite conclusion on the strength of the very same claim? Because, I believe, they disagree about what it takes to be a realist. Wood emphasizes that if the law is grounded in the nature of the will, then, more generally, it is grounded in the *nature* of something rather than any act I perform or attitude I have. And that, he believes, is sufficient, in the terms of contemporary metaethics, for Kant's position to be realism (Wood, 2008: p. 113). Korsgaard, on the other hand, appears not to believe this. She appears to believe that metaethical realism requires the constitution of moral facts, properties, or values independently of the will's nature, not just independently of any of its exercises.

3. The FA: Or, Systems of Heteronomy

In what remains I shall explain why, notwithstanding that Kant thinks the law is grounded in the nature of something—as Wood and Korsgaard both believe—it matters a great deal—contrary to Wood's judgment, and as Korsgaard would insist—which natures, and how many natures, we have in mind. I shall explain why this matters even if we reject the second Korsgaard's value-constructivist explanation of the will's function and, in general, any position according to which an act or attitude of the will plays an explanatory role in connection with the normativity of morality.[7] For in my view, the Rawlsians are right that there is a difference between the rationalists and Kant which realist readers underplay. Kant himself has a name for this difference. As he formulates it, while his rationalist predecessors endorse systems of *heteronomy*, Kant's ethics is a system of *autonomy*.

There are two major discussions of the autonomy/heteronomy contrast in Kant's published ethical works. One is in the *Groundwork*; the other is in the second *Critique*. In both discussions, Kant lays out a taxonomy of systems of heteronomy meant to include various ancient and modern moral psychologies. The principal division is between sentimentalist and rationalist conceptions of moral motivation. Conceptions of the sentimentalist kind he regards as obviously systems of heteronomy because, roughly speaking, according to them, I am morally motivated just insofar as there are some feelings with which I happen to find myself saddled. These feelings are not representations of anything normative for my action: they are not perceptions of the good or of duty. When I am motivated by them, therefore, I am not acting as I do *because* it is the right thing to do; I am merely chugging along in accordance with the idiosyncrasies of my psychological constitution. Rationalists may seem, on the other hand, to hold a view as different from this as any view could be. They say that I am motivated by my representations of the good, by an independently constituted normative order. When I am acting as I do, it is *precisely* because it is good to do so. And yet, as I shall show, Kant claims that this apparent difference between the sentimentalist and the rationalist is an illusion.

There are, I believe, several interconnected strands of argument for this claim to be found in the texts I have in mind, some of which I shall address in Section 5. For the moment, I wish merely to begin to state Kant's position with the aid of this exemplary passage:

[7]For a textual argument against such positions as readings of Kant, see Eckert-Kuang (2025).

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

> Wherever an object of the will has to be laid down as the basis for prescribing the rule that determines the will, there the rule is nothing other than heteronomy….Whether the object determines the will by means of inclination, as in the principle of one's own happiness, or by means of reason directed to objects of our possible volition in general, in the principle of perfection, the will never determines itself *immediately*, just by the representation of an action, but only by means of an incentive that the anticipated effect of the action has upon the will….For, because the impulse that the representation of an object possible through our powers is to exert on the will of the subject in accordance with his natural constitution belongs to the nature of the subject—whether to sensibility (inclination and taste) or to his understanding and reason, which by the special constitution of their nature employ themselves with delight upon an object—it would, strictly speaking, be nature that gives the law; and this, as a law of nature, must not only be cognized and proved by experience—and is therefore in itself contingent and unfit for an apodictic practical rule, such as moral rules must be—but it is *always only heteronomy* of the will… [*Grundlegung* 4:444][8]

In this passage, we are told that in a system of heteronomy, the special constitution of the "nature" of the subject, whether this arrangement is sensible or intellectual, means that its powers "employ themselves with delight upon an object." This suggests the following model for heteronomy: there's some object; I am naturally so constituted as to take delight in exercising my powers on it; for this reason, the object grounds laws for me. Characteristic of heteronomy is the explanatory priority of the object to my orientation towards it. More concretely, according to the *moral* heteronomist, there is something called *the good*, and I am so constituted as to want it and pursue it; I am, as we might put it, a lover of the good.

If the priority of the object to my orientation to it is characteristic of a system of heteronomy, then characteristic of a system of autonomy must be the opposite. It must be that, according to a system of autonomy, no explanatory significance is accorded to anything defined and constituted independently of the will, and in particular no explanatory significance is accorded to anything called "the good" defined independently of the will. And indeed, that is Kant's position. His notion of the good is precisely *not* the notion of something defined and constituted independently of the will, to which the will is then said to be oriented. Rather, the Kantian good *just is* the object of the will. By "the *good*" "is understood a necessary object of the faculty of desire" [*KpV* 5:58]; "the concept of good…must not be determined before the moral law…but only (as was done here) after it and by means of it" [*KpV* 5:63]. Our relation to the good is not a relation to a normative order lying outside our own nature.

And with that, I can give a preliminary statement of the positive part of my verdict on the Rawlsians. This idea, that Kant denies that our relation to the good is a relation to a normative order lying outside our own nature, is that in which Kant differs from his rationalist predecessors. I submit that the spirit of this idea is present in much work in the Rawlsian interpretive tradition and indeed in something that Rawls himself emphasizes both in his 1980 paper on Kantian constructivism and in his *Lectures*:

> Yet it suffices for heteronomy that [first] principles obtain in virtue of relations among objects the nature of which is not affected or determined by the conception of the person. Kant's idea of autonomy requires that there exist no such order of given objects determining the first principles of right and justice among free and equal persons. Heteronomy obtains not only when first principles are fixed by the special psychological constitution of human nature, as in Hume, but also when they are fixed by an order of universals or concepts grasped by rational intuition, as in Plato's realm of forms or Leibniz's hierarchy of perfections. (Rawls, 1980: p. 559).

[8]The corresponding passage in the second *Critique*—which, unlike the *Groundwork* passage, does not seem to allow the relevant aspects of our natural constitution to belong to our intellectual nature—is at *KpV* 5:41.

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

Of course, there is an important difference between the contrast I have drawn and the one Rawls here draws. Mine concerns what depends, and what does not depend, on our *nature*. His concerns are what depend and what do not depend on a *conception*: on the conception of the person in particular. But my point now is merely that the topic of this passage is the question whether we are to be taken to be connected to a normative order constituted independently of us. Its topic is not a procedure or the construction of value in valuing.

4. Two Forms of Constitutivism

But how exactly should we label this distinction between Kant and his rationalist predecessors? Rawlsians have traditionally called Kant a "constructivist," but—here I certainly agree with the realist readers—that really does not seem to fit. So how about one of the other terms that tradition has produced? I myself have indicated a preference for a constitutivist reading of Kant. So "constitutivist" seems, as a label, a good candidate.

But actually, the matter is more complicated than it may first appear. And that is because, though I indeed believe that Kant is a constitutivist, I do not believe that this is what distinguishes him from his rationalist predecessors, since I believe that at least some of them are constitutivists too. The distinction between Kant and his rationalist predecessors is one between two forms of constitutivism—a distinction that the more recent Rawlsians, who make use of the term "constitutivism," have not acknowledged. Indeed, Korsgaard claims that, as she puts it, the "combination of [a] rationalist view of normativity with a formal principle of reason is unusual" (Korsgaard, 2009: p. 64). But if by "formal principle" she means "internal standard" and by "rationalist view of normativity" she means an order of, as she sometimes puts it, "eternal normative verities" (*ibid.*), this combination is common enough, historically speaking. Korsgaard considers a position on which reason is "a normative module that has been inserted into you, for the purpose of making the laws of reason, which are essentially outside you, also be a force within you" (Korsgaard, 2009: p. 6). It is clear that she does not interpret this position as boasting the explanatory advantages of constitutivism. For it leaves open the skeptical question of why we should "conform to the demands of this little representative within us" (ibid.), whereas constitutive standards "meet skeptical challenges to their authority with ease" (Korsgaard, 2009: p. 29). This seems to be because constitutive standards apply to something by virtue of what it is, whereas the normative module, being merely inserted into you, is not what makes you what you are. But pre-Kantian heteronomist constitutivists believe precisely that your orientation to a normative order outside of you is (part of) what makes you what you are.

In particular, many of Kant's rationalist predecessors will agree that the standards of success in the exercise of the will are explained by the nature of the will: that is, that they are internal standards. This is different from another kind of position, according to which the nature of the will does not explain the standards for its exercise, but that there nonetheless are such standards. Good candidates for externalist readings include anyone with a purely elective conception of choice who endorses the doctrine of the liberty of indifference: Karl Leonhard Reinhold, for example, who famously complains that Kant is an externalist about choice (Reinhold, 1792: pp. 271–277). As I understand constitutivism, it is generic enough a notion that the differences between, say, Aquinas's more intellectualistic and Duns Scotus's more voluntaristic accounts of the relation of the will to the good reflect different implementations of constitutivism because in both cases, the good is explained in terms of the perfection of the will and thereby in terms of an internal standard.[9] Scotus, for example, writes that "the natural goodness of a volition…characterizes it insofar as it is a positive being—which indeed characterizes every positive being according to the degree of its being:

[9]For Scotus *contra* Aquinas on this point, see, e.g., *Ordinatio* I d. 1 part 2 q. 2 (Duns Scotus, 2017: 51–66). For helpful general discussion of the broader debate in the medieval period, see Hoffmann (2022).

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

the more being, the more goodness; the less being, the less goodness" [*Ordinatio* II d. 7 q. un., nn. 28–39 (Duns Scotus, 2017: p. 123)].[10] William of Ockham, on the other hand, may plausibly be read as endorsing no such implementation and as holding that there is no explanation, in terms of what the human will is, of the normative demand that it be related to the good. Thus, as Calvin Normore puts it, he may be read as divorcing the "action-guiding" sense of "good" from the "metaphysical" sense (Normore, 1998: p. 35). In general, if Ockham is to be read, as Normore suggests, as embracing an extirpation of internal teleology from created essences, then he is equally to be read as rejecting a constitutivist account of any normative standards applicable to creatures.

Constitutivists are known to complain that externalists cannot explain how external standards can be normative for the will, and this is a complaint that Kant would share. But it is not his complaint against his German rationalist predecessors, Leibniz and Wolff, because they are not externalists. They embrace internal teleology and follow figures like Aquinas and Duns Scotus in thinking that there is something in the nature of the will that orients it to the good and thereby constitutes the good as its internal standard. Leibniz, for example, writes that "the will is never prompted to action save by the representation of the good" [*Theodicy* 45] (Leibniz 1985) and, indeed, in commenting on the relation "of the will to the good," that "the effort to act in accordance with the judgment" of what is good "forms the essence of the will" [*Theodicy* 311 (Leibniz 1985)].[11] Wolffian perfectionism might be understood as a development of this Leibnizian idea: the will, for Wolff, is the capacity to promote perfection (e.g., Wolff (1720: pp. 10–11 *et passim*)).[12]

So Leibniz and Wolff, like Kant, can be read as constitutivists, as holding that exercises of the will are successful to the extent that they satisfy the will's internal standard. Nonetheless, they disagree with Kant about whether that is the end of the explanatory story. He says "yes," they say "no." Leibniz and Wolff say, in their different ways, that the will is *the capacity to pursue the good*, where "the good" is defined in normative terms independently of the will. Kant, in saying what the will is a capacity to do, does not use any independently normative notions. This is what distinguishes his *autonomist* constitutivism from their *heteronomist* constitutivism.[13,14]

[10]For some relevant passages, see *Quodlibet* I, q. 16 and IV, q. 1 (Ockham (1991: 75–6, 245–250)).

[11]It is a complex matter how the notion of perfection as applied to actions is related to the more general Leibnizian notion of metaphysical perfection as "the degree of positive reality, or what comes to the same thing, the degree of affirmative intelligibility, so that something more perfect is something in which more things worthy of observation are found" [Leibniz to Wolff, Winter 1714–15, GLW 161; Leibniz (1989: 230)], but like Aquinas and Scotus, the Leibnizians understand moral perfection as a kind of metaphysical perfection. For discussion, see Rutherford (2012). For discussion of the explanatory position of pleasure in relation to perfection for Leibniz, see Frey (2016). For discussion of Wolff, see Dyck (2024: 94–98) and Stratmann (2025).

[12]Kant reads Wolff as he reads Baumgarten, both as thinking of perfection as constituted prior to the will's functional normativity. This is why it makes sense for Kant to have been recorded as saying that "perfection and moral goodness are different" because "[p]erfection here"—i.e., in this case, in Baumgarten's sense—"is the completeness of man in regard to his powers, capacity, and readiness to carry out all the ends he may have," whereas "moral goodness consists in the perfection of the will, not the capacities" [*Collins* 27:265–66]. Kant's autonomist constitutivism says that it is the will's perfection alone which matters normatively to the will, not in the sense that it is the notion of something with independent normative weight to which the will is directed, but merely in the sense that it is the same as the will's successful exercise (or the state that conduces to it).

[13]This means that my reconstructions in §2 of Wood's and the second (but not the first) Korsgaard's positions are both examples of heteronomist constitutivism. Since the second Korsgaard is a constructivist, heteronomist constitutivism need not be realism.

[14]I use "constitutivist about X" as a label for those who believe that a normative standard applies to X either because it is part of what constitutes, or part of what springs from, the essence or nature of X. (For discussion of the difference between these disjuncts, see Fix (2018).) I use "externalist about X" as a label for those who reject constitutivism about X but nonetheless believe that normative standards apply to X. There is a real division, both historically and philosophically speaking, between constitutivists and externalists, so described. In particular, externalism about X contains on its own no explanation of the connection of X to the normative standards that govern it, whereas constitutivism about X, according to either disjunct, is just such an explanation, and in both cases, one in which X's nature plays the starring role. Of course, there are other determinate explanatory *desiderata*—like some of those I discuss below, in the discussion of Kant's arguments—which are satisfied by autonomist, but not heteronomist, constitutivism, and many philosophers interested in constitutivism might be moved

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

And with that in hand, I can state my final verdict on the Rawlsians. What they have right is not that the rationalist thinks that the moral law is grounded in the nature of things, whereas Kant thinks that it's grounded in some act or attitude. Kant agrees with the rationalists that it's grounded in the nature of something. Nor even is it that while Kant thinks the law is grounded in the nature of the will, the rationalist thinks it's grounded in the nature of something else. Leibniz and Wolff agree with Kant and disagree with figures like Ockham on this point. What they have right is that, for Kant, unlike for the rationalist, there is no normative order independent of the nature of the will in terms of which its internal standard is to be understood in turn. For Kant, the explanatory buck stops with the nature of the will in a way in which it does not for the rationalist and her medieval antecedents.

5. Kant's Arguments

At this point, I claim to have done what I promised, which was to say what is right and what is wrong in the Rawlsian reading, or at least in its spirit. But one thing I have not done is to consider Kant's arguments for autonomist constitutivism. This section will include the sparest sketch of some of these arguments, intended not to eventuate in a judgment in his favor but rather merely to sharpen the description of his position.

In the passage from the second *Groundwork* considered above, there are two reasons given for Kant's position: first, that only autonomist constitutivism accounts for the *a priori* character of our knowledge of the good; second, that only autonomist constitutivism accounts for the categorical character of moral requirement. These two reasons are, moreover, easy to construe as a package deal, because he thinks that moral requirement can be categorical *only if* our knowledge of the good is *a priori* [e.g., *Grundlegung* 4:389 f.]. I myself doubt that he can make a convincing case for that claim or that he can show that none of his predecessors can account for categoricity.[15] His basic thought is that for them, moral principles can only be hypothetical imperatives, because they instruct the agent to do something to which she is oriented by default. But I suspect that this criticism betrays a lack of appreciation for the structural complexity of the positions to which he opposes his own. To take one example, Aquinas says that in its default orientation, the will is directed to the good of its possessor; it is only when it possesses the virtue of justice that it is equally directed to the good of others [*DQV* 5c; Aquinas (2014: pp. 522–523)]. So moral commandments are *not* hypothetical imperatives relative to the agent's default orientation. Of course, what happens when the will has its virtue will be described by Aquinas and Kant in very different ways, but that is a further issue. That a notion with the structural features of obligation can be constructed for a heteronomist constitutivist, along the lines I have suggested, seems tolerably clear. If it cannot, Kant himself does not do the work of showing this.[16]

Thus, I take leave to set aside the issue of categoricity and to ask whether independent arguments can be given which focus on apriority. I begin by operating at one remove from the text, to which I shall return presently. So here's a Kant-style argument to the effect that the heteronomist cannot account for the apriority of our knowledge of the good. The heteronomist says that we human beings are lovers of the good, which is something constituted and defined independently of us. But how does she know that? To know that, you have to know three things: human beings, the good, and a relation between them. But how does she know the good? *Ex hypothesi*, not empirically. So does she have an intellectual intuition of the good? Or is this rather to be understood as a theory that we should accept on the strength of its various virtues by application of inference to the best explanation?

primarily by these or similar *desiderata*, in which case heteronomist constitutivism is of little use to them. But its difference from externalism and its commonalities with its autonomist sibling are no less real for that.

[15]This is the principal contention of Silber (1960).

[16]For a detailed and illuminating defense of Aquinas against Kant on this point, see Hinton (2002).

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

You might like some of these epistemological alternatives. Then again, like Kant, you might not. You might think that the idea that we are possessed of intellectual intuition is enthusiasm akin to dreams of a spirit seer. You might think that assessing theories according to their virtues has no place in philosophy, in which hypotheses are contraband [Axv]. You might agree with Kant, on the positive side, that to the extent that we have *a priori* knowledge of something, that something cannot be constituted independently of us.

This essay is not, of course, the venue for a comprehensive examination of Kant's arguments for this positive claim about *a priori* knowledge. But it will help in the present context to draw out some of the attractions of this position on the *a priori* in application to the case of morally motivated action.[17] This will also allow me to get to the explanation I have promised of the sense in which the sentimentalist and the rationalist are both perspicuously called heteronomists, notwithstanding that their positions initially appear to be as different from one another as any two positions on the same issue might be, the sentimentalist asserting that I do what I do on account of the idiosyncrasies of my constitution and the rationalist asserting that I do what I do in observance of an external normative order.

Recall that the heteronomist characterizes the will in the independent terms of the good: the will is the capacity to pursue the good. I exercise the will successfully to the extent that I do so in pursuit of the good. But now let us consider another theorist of the will, one who takes issue with our heteronomist's characterization of its nature. Our new theorist says that the will is a capacity to pursue not the good but the gluttonous consumption of biscuits. She is not an *agathist* but a *bisculinarist*. The bisculinarist says that to the extent that I pursue the gluttonous consumption of biscuits in the exercise of the will, I exercise it successfully; and that to the extent that I do not, I exercise it unsuccessfully.

Doubtless this conception of the human will makes for an image that is ridiculous rather than sublime. But notice that, as far as the constitutivist explanatory machinery is concerned, it works in exactly the same way. Success and failure of exercise are determined by the nature of the capacity; it matters not whether the nature of the capacity is spelled out in an infinitival expression that makes use of independently normative vocabulary. That you succeed is just a matter of satisfying the internal standard of the capacity you exercise; that you fail is just a matter of not satisfying it.

This observation is what leads us to the problem with heteronomist constitutivism, from Kant's point of view, but to see how, it will help to taxonomize one more time, and divide the sphere of heteronomist constitutivism in two: for the argument must be run differently for each of these two varieties. The heteronomist constitutivist says that the will is the capacity to pursue the good, where the notion of the good is an independently normative notion. This entails that the will is the capacity to φ, where φing is equivalent to pursuing the good. But now we may distinguish between a heteronomist who says that the nature of the will itself takes account of this equivalence—so that the will is oriented towards the good *as such*—and a heteronomist who says that it does not, so that there is a mere harmony of the will and the good, though no direct explanatory relationship between them. There is no difficulty in characterizing a sense in which the latter heteronomist cannot claim that the acts of our will observe an external normative order. According to her, when I exercise my will well, I φ because φing constitutes success in such exercise; that φing is equivalent to pursuing the good is, by her own admission, neither here nor there.

It takes a more extended argument to characterize a sense in which the former heteronomist—who is more to the point, being the type of those of Kant's predecessors I have chiefly in mind—cannot claim that the acts of our will observe an external normative order. First we need a more detailed characterization of some features of the will. The will is a desiderative or appetitive power, a power of knowledge the object of whose act is existentially dependent on the act itself. It is also a self-conscious capacity, in the sense that it admits of a kind of exercise that contains and is explained by

[17]For an analogous discussion in the theoretical case, see Pendlebury (2022: pp. 381–2).

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

its bearer's knowledge of that exercise as successful: that is, as meeting the capacity's internal standard. For the bearer to know, internally to the exercise of the will, that this exercise meets the will's internal standard is *a fortiori* for the bearer to know, internally to the exercise, the standard itself and thereby the nature of the will as a standard-governed capacity. In such cases, then, the will's act is knowledge both of what depends existentially on this act and of something—the standard of exercise—which is prior to the act but not prior to the nature of the will.

And this, moreover, exhausts what can be known internally to the exercise of the will. For apart from what its exercises bring into being and its own nature, all that there is to know is that which is prior both to the will's acts and to its nature. But the will cannot be a capacity to know any such thing, for it is not a cognitive or apprehensive power but a desiderative or appetitive power. And this is what entails that there is a sense in which the heteronomist, even if she embraces a direct explanatory dependence of the will's nature on a normative order outside it, cannot claim that its *exercises* can observe this normative order. She cannot claim, namely, that these exercises contain knowledge of themselves as in agreement with this normative order as such, because they cannot contain knowledge of that order as such. Thus, even for this kind of heteronomist, when I exercise my will well, though in one sense I do what I do because an external normative order demands it, no knowledge of this explanatory relation is internal to my act. Insofar as what explains my act is knowledge internal to the exercise of the will, it is knowledge merely of the conformity of the exercise to the will's standard. That the latter is explained by an external normative order is neither here nor there. The significance of the difference between the sublimity of agathism and the ridiculousness of bisculinarism escapes the self-consciousness of my act of volition as such.

And this is the sense in which the rationalist and the sentimentalist are, from Kant's point of view, of a piece. The sentimentalist says that I do what I do not because of the demands of an external normative order but merely on account of the contingencies of my constitution. The heteronomist constitutivist wants to say that I do what I do in recognition of an external normative order, but the structure of her position means that she is forced to agree with the sentimentalist. Sentimentalism and constitutivist rationalism both have the consequence that I do what I do, *not* because it is (externally) good, but—insofar as this is a self-conscious "because" internal to the exercise of the will itself—merely on account of the contingencies of my constitution. This is the real reason why, as Rawls repeatedly insists, Kant rejects moral rationalism "as a form of heteronomy just as firmly" as he rejects "Hume's psychological naturalism" (Rawls, 2000: p. 235).

Now, having introduced that argument at two removes from the text, I want to return to the text —but not to the text of any of Kant's practical works. I want to consider something in the *Critique of Pure Reason*, something that he says about the categories of the understanding. At the point at which he makes this remark, he purports to have shown that the categories and concepts like *reality* and *negation*, *substance* and *cause*, express ways in which we must, as he puts it in the *Prolegomena to Any Future Metaphysics*, "spell out" [*buchstabieren*] the structure of perceived objects, lest we lose our grip on their *being* perceived objects: that is, *genuinely* perceived, rather than imagined [*Prolegomena* 4:312]. He has argued that the categories are, moreover, *no more than this*: that their significance is *exhausted* by this role that they play in spelling out the structure of the perceptible. In saying this, he takes himself to disagree with those of his predecessors who believe that these concepts refer to something constituted independently of the human intellect, something to which, in exercising the intellect, we might hope to be attuned. Here is what he says about one version of that view, a version he calls a "preformation system":

> If someone still wanted to propose a middle way between the only two, already named ways, namely, that the categories were neither **self-thought** *a priori* first principles of our cognition nor drawn from experience, but were rather subjective predispositions for thinking, implanted in us along with our existence by our author in such a way that their use would agree exactly with the laws of nature along which experience runs (a kind of **preformation-system** of pure reason), then…this would be decisive against the supposed middle way: that in

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

> such a case the categories would lack the **necessity** that is essential to their concept. For, e.g., the concept of cause, which asserts the necessity of a consequent under a presupposed condition, would be false if it rested only on a subjective necessity, arbitrarily implanted in us, of combining certain empirical representations according to such a rule of relation. I would not be able to say that the effect is combined with the cause in the object (i.e., necessarily), but only that I am so constituted that I cannot think of this representation otherwise than so connected… [B167-8]

Note the parallels between this and heteronomist constitutivism. There is meant to be something that is there, independently of us—here, the "laws of nature," there, the good—and we are predisposed by virtue of the constitution of our subject to think, in the theoretical case, or act, in the practical case, in accordance with it. In this passage, Kant says that this notion of accord, in the theoretical case, is a sham. The preformationist wants it to lie in our nature that we are connected to an order of causal intelligibility outside of ourselves, by virtue of a disposition that we have to think in a certain way, a way that matches up with that order of intelligibility. The trouble is that, on the preformationist's picture, even if this is true, that this is an order of intelligibility is not reflected in our cognitive acts themselves. When I judge that the sun warms the stone, I do so not *because* the sun warms the stone (even if that is true), but because I am disposed to make such a judgment in my circumstances. Thus, as soon as the disposition is introduced to the explanatory story, the external order of intelligibility is bleached of its explanatory import in connection with my judgment. This is why Kant says, even under the assumption of that external order and a match-up between our disposition and it, that when I exercise this disposition in applying the concept of *cause*, I do not count as grasping the causal relation that is really there; I'm just chugging along as I'm disposed to.

Not every rationalist is a preformationist about the intellect, of course; here Kant has Crusius in particular in mind, as is explicit at *Prolegomena* 4:319n. A rational intuitionist, for example, would say not that the intellect has implanted dispositions to judge in such-and-such a way but that we have non-sensible modes of access to objects of knowledge. But the heteronomist constitutivist is analogous to the preformationist, and not to the rational intuitionist, in the sense I have indicated. This is true even of the heteronomist who embraces a genuine explanatory connection between the will and the good. This is precisely because, as I have emphasized, the will, unlike the theoretical intellect, is an appetitive or desiderative power: it is not a power to know anything prior to itself. This is what makes rational intuitionism unavailable in the practical case. This means that even a heteronomist who holds that the good is implicated in the nature of the will as a capacity must admit that in an exercise of the will, I do not know my exercise to be explained by the good.

This is not to say, of course, that the heteronomist must accept that we live in total ignorance of the good and its explanatory connection to acts of the will. On the contrary, she can, and typically does, say that we know the good and its relation to the will and the latter's exercise in an exercise of the intellect, and even that such an exercise of the intellect can itself explain the exercise of the will.[18] But what Kant can say, which the heteronomist cannot, is that there is nothing to the good and what it is to pursue it that escapes the exercise of rational appetite itself: nothing to the good that escapes the self-consciousness of volitional activity.

To say this is not, of course, to offer an argument to the heteronomist with the ambition of immanent conviction. My aim has been to reach a more precise description of Kant's disagreement with the heteronomist rather than a judgment in his favor. Kant's alternative will, of course, be attractive to those who share his general position on *a priori* knowledge, to the effect that we cannot know *a priori* anything constituted independently of our powers of knowledge, especially if they also believe, as he does, that the good is not constituted by the nature of the theoretical intellect,

[18]Consider, for example, Aquinas: "with respect to the determination of an act, which depends on the object, the intellect moves the will, since what is good is also apprehended according to a particular notion that is contained in the universal notion of what is true" [*Summa theologiae* 1a2ae 9.1 ad3 (Aquinas, 2014: p. 400)].

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

since in that case the only thing left for it to be constituted by is the will. But even if we set aside Kant's general position on the *a priori*, this section's consideration of Kant's arguments has added something to the previous section's statement of his position: namely, that heteronomist constitutivism not only makes the good constitutively external to the will but also, because the will is an appetitive power, makes the good as such epistemically external to the will's exercise. It may be that there are reasons apart from Kant's general position on the *a priori* to object to the latter implication.[19]

6. Objections and Replies

Before I conclude, I want to consider two instructive Kant-internal objections to the argument against heteronomist constitutivism that I have just presented. First, there is an objection to Kant's anti-preformationist argument, which occurs to many readers of the first *Critique*, and an analogous objection arises in the practical case. Kant himself, like Hume, denies that in making causal judgments, I am attuned to a causal order constituted independently of my cognition. So has he not merely joined the empiricists against the rationalists? Likewise, Kant himself, like Hume, denies that in the exercise of the will, I am attuned to a normative order constituted independently of it. So must he not agree with Hume that I do what I do not because it is good but merely because I am disposed to do it?

No.[20] For Kant, there is no notion of the good defined independently of the will. And this makes him fundamentally different from the sentimentalist. Sentimentalism concedes to rationalism that there is something which it would be for us to be connected to an external normative order; it just denies that we are thus connected. The very articulation of the sentimentalist position, as a skeptical moral-psychological position, depends on the availability of the notion of such a connection. That may seem dubious, inasmuch as Hume denies that we have ideas of right and wrong [*Treatise* III.i.i (Hume, 1978: pp. 467–9)]. But we do, for Hume—at least as Kant reads him—have a sense of what the ideas of right and wrong would have to be if we had them. He says that though we thought our moral thought and talk were about an external normative order, they are not: actually, they're just an expression of feeling. If there is a normative order in things, it is not one that we can know or have any idea of. This is parallel to his skeptical empiricism in the theoretical case. There he says, we can know no causal relations in things; causal judgments are just the expression of feelings (in that case, of expectation developed in reaction to constant conjunction). If there is a causal relation in things, it is not one that we can know or have any idea of. The articulation of these positions depends on a distinction between things, unstained by the palette of our minds, and such painterly embellishments as we bestow upon them. The rationalist and Hume disagree about how much we can justifiably say about what is in the things as opposed to what we project on them.[21]

Kant's anti-skeptical move, which is equally his anti-heteronomist move, is to deny the availability of the notion of things independent of us and our palette required to articulate both the dogmatic rationalist and the skeptical empiricist positions.[22] The skeptic can make us look parochial only if she can compare what we say and feel about things with what they, independently of that, really are; we can say to what extent our representations of the things conform to them and

[19]See, for example, for arguments against what he calls "intellectualism," Fix (2018).

[20]Sentimentalism is not typically developed so as to include internal-standard normativity as a basic explanatory component. Normative notions are typically constructed from non-normative materials. But this complication is not germane to the present discussion.

[21]I here prescind from the issue of how things are meant to be related to impressions for Hume. I am understanding him as Kant understands him: as thinking that impressions disclosed to us features of things in themselves.

[22]Kant's notion of things in themselves is not such a notion because things as they are in themselves are not cognitively available to us to stain with our mental palette; Hume's skepticism depends essentially on such cognitive availability through impressions. Thanks to Chandler Hatch for this clarification.

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

to what extent they do not. But Kant does not permit any such comparison, since for Kant, it is not our *a priori* representations that conform to things, but the things that conform to our *a priori* representations. His Copernican revolution, in the practical as in the theoretical case, is as antiskeptical as it is antidogmatic: for it is the rejection of the dogmatic assumption upon which skepticism depends just as much as dogmatism does, that in philosophy we may help ourselves to a general notion of things, the real, the world, and what's out there, understood as a notion intelligible independently of any notion we may have of our own intellect and will.[23]

But there is a second objection, against which Kant's Copernicanism may seem, at first, to be of no use at all, and that is that it leaves open that the constitution of our intellect and will may be ridiculous rather than sublime. In particular, it may seem that it leaves open that bisculinarism is a perfectly good description of our moral nature, alongside Kant's preferred description in terms of the moral law. After all, the bisculinarist seems to count as an autonomist constitutivist, since she denies that the internal standard of the will is grounded in a normative order outside its nature.[24] And if so, can we not after all generate a sense in which our nature is idiosyncratic, and thereby see Kantianism collapse into something like sentimentalism?

This is a topic that I cannot treat adequately here.[25] Kant himself would want to exclude bisculinarism by arguing that it is unintelligible that anyone might be moved to pursue the gluttonous consumption of biscuits apart from happening to be inclined towards it; thus, he would say, bisculinarism does not offer an intelligible principle of fundamental end-selection. Only the moral law does. I myself am not sure whether he can explain why, independently of its being fully intelligible to us, constituted as we are, that someone might do something solely and fundamentally because she owes it to someone else, even if she does not want to do it; whereas it is not intelligible to us, constituted as we are, that something might do something solely and fundamentally because it would allow her to eat biscuits, even if she does not want to eat biscuits.

It *may* be that we can say more than this. It might be suggested that further conditions of adequacy on principles of fundamental end-selection rule out at least some competitors to the moral law: for example, that the principle be serviceable as that of nature, or even as that of a living nature.[26] And certainly Kant seems to make such suggestions himself, and they may be correct. But even if they are not, I suspect that he can make good on his claim that the moral law is the only principle of fundamental end-selection whose intelligibility we are in a position to appreciate, and that this is enough to avoid a judgment of idiosyncrasy akin to the one characteristic of sentimentalism. This is because the mere logical possibility of bisculinarism (if it is logically possible) does not make it a really intelligible competitor to Kant's conception of the will—that is, a real possibility that we can know and understand—any more than the mere logical possibility of forms of intuition different from our own make them really intelligible competitors to space and time.[27] To judge something to be an idiosyncrasy, we must compare it with something else and see that it is different. But bisculinarism is not a something, not a reality, to which *we* can compare anything; it is, as far as we know, no more than a logically consistent description, in which we can understand no reality, nature, essence, or form, since a bisculinarist is, to us, completely unintelligible as a motivational and normative possibility.

7. Conclusion

This last observation leads naturally to one last description of the difference between Kant and his predecessors, one which brings out the connection of the themes of this essay to the broader significance of Kantianism as a kind of philosophy.

[23]Thanks to Mathis Koschel for the insight that this is essential to Hume's skepticism about the idea of necessary connection, which was crucial in the development of this line of argument.

[24]Thanks to Bennett Eckert-Kuang and Banafsheh Beizaei for pressing me on this question.

[25]But see Pendlebury (2024) for an account of the methodology of the *Groundwork*, which bears on this question.

[26]Thanks to Barbara Herman and Chandler Hatch for these thoughts.

[27]See Gurofsky (2020) for an argument for this latter claim.

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

Rawlsians have been accused of reading Kant as trying to "cut morality down" to the human domain, to use Karl Ameriks's description (Ameriks, 2003: p. 271). This may be a perspicuous description of the ambitions of some contemporary philosophers who find inspiration in Kant, some Rawlsians among them. Certainly this is suggested when, say, constructivists are classed with expressivists as against realists. But what I said in the previous section, by way of contrasting Kant with Hume, means that my reading cannot be included in Ameriks's target range. To cut morality down to the human domain, we require a contrast of the human domain with something else: we have to say that morality is not out there, merely in here. That is not what Kant is doing when he defines the good in terms of the finite rational will.

To see why, it helps to place Kant's autonomist constitutivism in the context of his broader Copernican turn. For the opposition of autonomist to heteronomist constitutivism is no isolated incident: it is part of Kant's broader displacement of one kind of philosophy and, correlatively, one kind of philosophical anthropology, with another. According to the traditional kind, the subject matter of philosophy is marked out not merely by the good, but by a suite of convertible "transcendentals," which includes, apart from the good, being, one, and the true. The human being is, in turn, depicted as the earthly being who enjoys the privilege and distinction of relating to these, the hallmarks of philosophy, in a particular way. The human intellect is trained on the true, and the human will on the good, both, thereby, on being itself. To "cut morality down" to the human domain would be to take the traditional image on its own terms, remove the good, and claim that morality remains, though localized in the human being, with a blank space where the good used to be, a space that only appears to be occupied when the human being projects occupants into it.

That is not what Kant does. For Kant agrees with the traditionalist that the subject matter of philosophy is marked out by being, one, the true, and the good; that they fill the space depicted by philosophy without remainder. But he disagrees on the second point, about the human being. Kant's Copernican revolution, which has it that the objects conform to us and not we to them, means that the true is, as such, the object of our intellect, and that the good is, as such, the object of our will. Thus, it's not that the human intellect and will are of interest in philosophy because the true and the good are. It's the other way round: the latter are hallmarks of philosophy because, and only because, philosophy is, as such, all about us.

Thus, from Kant's point of view, though the tradition is right that the good belongs to the will, it's wrong about why. The good is ours not on account of any privilege and distinction that we may enjoy among earthly beings. Rather, it belongs to us because we belong to it. The objects conform to our cognition, in the practical as well as in the theoretical case, in the sense that, in philosophy, they are defined in terms of us.

Acknowledgments. I wish to thank, for comments and conversation, Lucy Allais, Banafsheh Beizei, Anastasia Berg, Janelle DeWitt, Kyla Ebels-Duggan, Bennett Eckert-Kuang, Sonny Elizondo, Stephen Engstrom, Luca Ferrero, Jeremy David Fix, Sare Ghorbani Khaledi, Matthias Haase, Rafeeq Hasan, Chandler Hatch, Barbara Herman, David Horst, Ben Jia, Mikayla Kelley, Christine Korsgaard, Mathis Koschel, Jed Lewinsohn, Jürgen Lipps, Alexandra Newton, Kate Nolfi, Calvin Normore, Hille Paakkunainen, Nataliya Palatnik, Japa Pallikkathayil, Jordan Pascoe, Michael Pendlebury, Daniel Ranweiler, Karl Schafer, Keshav Singh, Sergio Tenenbaum, Mary Tjiattas, and audiences at the conference on Constitutivism in Ethics and Epistemology at the University of Lisbon; the Practice and Theory conference at the University of California, Los Angeles; the 2026 Royal Conference at the University of Texas at Austin; the conference on Kantian Moral Constitutivism at the University of Chicago; and a meeting of the German Philosophy Workshop at the University of Chicago.

T. A. Pendlebury is Assistant Professor of Philosophy at the University of Chicago. His principal research interests are in the theoretical and practical philosophy of Kant and its connections to early modern philosophy and German idealism.

References

Works by Kant

Citations of Kant's works (except for the *Critique of Pure Reason*, in citing which I supply A/B pagination) are to the Academy Edition. Translations are based on those of the *Cambridge Edition of the Works of Immanuel Kant* unless otherwise noted.

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

[*A/B*]	(1781/7) *Kritik der reinen Vernunft*
[*Prolegomena*]	(1783) *Prolegomena zu einer jeden künftigen Metaphysik die als Wissenschaft wird auftreten können*
[*Collins*]	(1784–5) *Moralphilosophie Collins*
[*Grundlegung*]	(1785) *Grundlegung zur Metaphysik der Sitten*
[*KpV*]	(1788) *Kritik der praktischen Vernunft*

Works by Others

Ameriks, K. (2003). *Interpreting Kant's Critiques*. Oxford University PressOxford. https://doi.org/10.1093/0199247315.001.0001

Aquinas, T.J. Hause & R. Pasnau (Eds.). (2014). *Thomas Aquinas: Basic works*. Hackett Publishing Company.

Duns Scotus, J.T. Wiliams (Ed.). (2017). *John Duns Scotus: Selected writings on ethics*. Oxford University Press.

Dyck, C. W. (2024). *Wolff and the first fifty years of German metaphysics*. Oxford: Oxford University Press.

Eckert-Kuang, B. (2025). Kant on self-legislation as the foundation of duty. *European Journal of Philosophy*, 33(3), 910–926. https://doi.org/10.1111/ejop.13026

Fix, J. D. (2018). Intellectual isolation. *Mind*, 127(506), 491–520. https://doi.org/10.1093/mind/fzx046

Fix, J. D. (2021). Two sorts of constitutivism. *Analytic Philosophy*, 62(1), 1–20. https://doi.org/10.1111/phib.12166

Fix, J. D., & Pendlebury, T. A. (Forthcoming). The rational faculty of desire. In S. Bacin & C. Bagnoli (Eds.), *Reason, agency, and ethics: New perspectives on Kantian constitutivism*. Oxford University Press.

Frey, J. (2016). Was Leibniz an Egoist? *Journal of the history of philosophy*, 54(4), 601–624. https://doi.org/10.1353/hph.2016.0072

Gurofsky, S. (2020). On the putative possibility of non-spatio-temporal forms of sensibility in Kant. *European Journal of Philosophy*, 28(4), 841–856. https://doi.org/10.1111/ejop.12536

Hinton, T. (2002). Kant and Aquinas on the priority of the good. *The Review of Metaphysics*, 55(4), 825–846.

Hoffmann, T. (2022). *Free will and the rebel angels in medieval philosophy*. Cambridge University Press.

Hume, D. (1978). *A treatise of human nature*. Oxford University Press.

Kain, P. (2017). Dignity and the paradox of method. In R. dos Santos & E. E. Schmidt (Eds.), *Realism and antirealism in Kant's moral philosophy* (pp. 67–90). De Gruyter.

Kleingeld, P., & Willaschek, M. (2019). Autonomy without paradox: Kant, self-legislation and the moral law. *Philosophers' Imprint*, 19(6), 1–18.

Korsgaard, C. M. (2009). *Self-constitution: agency, identity, and integrity*. Oxford University Press.

Korsgaard, C. M. (1996a). *Creating the kingdom of Ends*. Cambridge University Press.

Korsgaard, C. M., (1996b). *The sources of normativity*. Cambridge University Press.

Leibniz, G. W. (1985). *Theodicy*. Translated by E. M. Huggard. Open Court.

Leibniz, G. W. (1989). *Philosophical essays*. Translated and edited by Roger Ariew, & Daniel Garber. Hackett Publishing Company.

Normore, C. G. (1998). Picking and choosing: Anselm and Ockham on choice. *Vivarium*, 36(1), 23–39. https://doi.org/10.1163/1568534982579378

O'Neill, O. (2012). *Constructions of reason: Explorations of Kant's Practical Philosophy*. Cambridge University Press.

Ockham, William. (1991). *Quodlibetal questions*. Translated by A. J. Freddoso, & F. E. Kelley. Yale University Press.

Pendlebury, T. A. (2022). The shape of the Kantian mind. *Philosophy and Phenomenological Research*, 104(2), 364–387. https://doi.org/10.1111/phpr.12767

Pendlebury, T. A. (2024). The will of all in *Kant's groundwork*. *Kantian Review*, 29(3), 423–445. https://doi.org/10.1017/S1369415424000347

Pendlebury, T. A. (2025). The Kantian's Desire. *Philosophy and Phenomenological Research*, 111(2), 652–675. https://doi.org/10.1111/phpr.70041

Rawls J., B. Herman (Ed.). (2000). *Lectures on the history of moral philosophy*. Harvard University Press.

Rawls, J. (1980). Kantian constructivism in moral theory. *Journal of Philosophy*, 77(9), 515–572. https://doi.org/10.2307/2025790

Reath, A. (2006). *Agency and Autonomy in Kant's Moral Theory*. Oxford University PressOxford.

Reinhold, K. L. (1792). *Briefe über die Kantische philosophie, Zweyter Band*. Georg Joachim Göschen.

Rutherford, D. (2012). *Leibniz and the rational order of nature*. Cambridge University Press.

Schafer, K. (2015). Realism and constructivism in Kantian metaethics (1): Realism and constructivism in a Kantian context. *Philosophy Compass*, 10(10), 690–701. https://doi.org/10.1111/phc3.12253

Schafer, K. (2019). Kant: constitutivism as capacities-first philosophy. *Philosophical Explorations*, 22(2), 177–193. https://doi.org/10.1080/13869795.2019.1599049

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press

Schafer, K. (2021). Transcendental philosophy as capacities-first philosophy. *Philosophy and Phenomenological Research*, 103(3), 661-686. 675. https://doi.org/10.1111/phpr.12740

Sensen, O. (2018). Kant's constitutivism. In R. dos Santos & E. E. Schmidt (Eds.), *Realism and Antirealism in Kant's moral philosophy* (pp. 197–220). De Gruyter.

Silber, J. (1960). The Copernican revolution in ethics: The good reexamined. *Kant-Studien*, 51(1–4), 85–101. https://doi.org/10.1515/kant.1960.51.1-4.85

Stratmann, J. (2025). Autonomy without compromise: Wolff, Kant, and the grounds of moral laws. *Journal of the History of Philosophy*, 63(1), 97–120. https://doi.org/10.1353/hph.2025.a950465

Wolff, C. (1720). *Vernünftige Gedancken von der Menschen Thun und Lassen*. Renger.

Wood, A. W. (2008). *Kantian ethics*. Cambridge University Press.

Cite this article: Pendlebury, T. A. 2025. The Will and the Good. *Canadian Journal of Philosophy* 55: 560–576, doi:10.1017/can.2026.10090

https://doi.org/10.1017/can.2026.10090 Published online by Cambridge University Press