



Autonomy and Objective Moral Constructivism: Rawls Versus Kleingeld & Willaschek

Alyssa Rose Bernstein¹

Received: 12 January 2023 / Revised: 12 January 2023 / Accepted: 19 January 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

Pauline Kleingeld and Marcus Willaschek, in a co-authored article, declare that their purportedly new interpretation of Immanuel Kant's writings on autonomy reveals that his moral philosophy is neither realist nor constructivist. However, as I explain here, John Rawls already occupies the area of intellectual territory to which Kleingeld and Willaschek attempt to lay claim: Rawls interprets Kant's moral philosophy as neither realist, as Kleingeld and Willaschek evidently construe this term, nor constructivist, as they evidently construe this term. Contra Kleingeld and Willaschek, the moral constructivism attributed to Kant by Rawls is not voluntarist, and Rawls's account of Kant's concept of autonomy is not paradoxical. In order to understand autonomy, it is necessary to understand Kant's complex conception of the will, which structures his moral philosophy (as Rawls, unlike Kleingeld and Willaschek, explains). Rawls, like Kant, but unlike Kleingeld and Willaschek, clearly distinguishes between certain importantly different questions about normativity and obligation. Kant's moral philosophy, according to Rawls's insightful interpretation, is a form of objective moral constructivism.

Keywords Autonomy · Constructivism · Immanuel Kant · Moral law · Moral objectivity · Moral obligation · Normativity · John Rawls

1 Introduction

Autonomy, the concept central to Kant's moral philosophy, is disputed among its interpreters. Some offer paradoxical accounts of it (whether consciously or not), which Kant's defenders critique. Here I defend John Rawls's interpretation of Kant's moral philosophy from a charge recently issued by Pauline Kleingeld and Marcus Willaschek, who claim that Rawls's account of autonomy is paradoxical.

✉ Alyssa Rose Bernstein
bernstei@ohio.edu

¹ Philosophy Department, Ohio University, 1 Ohio University, Athens, OH 45701, USA

In defending it I argue that Rawls (unlike Kleingeld and Willaschek) clearly distinguishes between certain importantly different questions about normativity and obligation, as does Kant. I also argue, contra Kleingeld and Willaschek, that the moral constructivism attributed to Kant by Rawls is not voluntarist.¹

Much is at stake in these debates about moral constructivism and autonomy. Justly assessing Rawls's interpretation of Kant's moral constructivism is necessary for justly assessing Rawls's own political constructivism.² Some who write about Rawls's philosophical work criticize it incorrectly, due to their having failed to distinguish clearly between Rawls's political constructivism and his interpretation of Kant's moral constructivism.³ Also at stake in these scholarly debates, as I show below, is the cogency of Kant's philosophical conception of reason.

Kleingeld and Willaschek, in their co-authored article, "Autonomy Without Paradox: Kant, Self-Legislation and the Moral Law"⁴ (hereinafter, "K&W"),⁵ survey and critically examine accounts of autonomy offered by many scholars (some of whom interpret Kant as a moral realist, others as a moral constructivist), and argue that those who interpret Kant as a moral constructivist incorrectly ascribe to him the following set of paradoxical claims: (a) to be a morally autonomous person, or to have a morally autonomous will, is to give the Moral Law to oneself, or to self-legislate it;⁶ (b) this act of law-giving or self-legislation is the source of the normativity of the Moral Law; and (c) the Moral Law is universally and unconditionally normative. This set of claims is paradoxical because, taken together, they characterize the Moral Law as both unconditionally normative and conditionally normative. I will refer to this set of claims as "the paradox of self-legislation."

¹ Allen W. Wood argues that the moral constructivism attributed to Kant by Rawls is voluntarist, subjectivist and/or relativist. See his book, *Kantian Ethics* (Cambridge University Press, 2008; hereinafter, "KE"). Wood makes such arguments throughout this book; see, for example, pp. 46, 51–54, 282–285. See also Wood's book, *Formulas of the Moral Law: Elements in the Philosophy of Immanuel Kant* (Cambridge University Press, 2017), p. 2. I do not attempt here to reply to Wood's arguments, since this requires too many pages; I do so in a manuscript in progress, in which I argue that although Wood misinterprets Rawls's writings, there is a high degree of similarity between Wood's and Rawls's interpretations of Kant.

Disclosure: Wood was one of my warmly esteemed professors while I was an undergraduate student at Cornell University, and he has been and continues to be a highly valued mentor; Rawls was one of my warmly esteemed professors and PhD dissertation advisors while I was a graduate student at Harvard University; I worked for him as a teaching assistant and a research assistant, and he was a highly valued mentor.

² I discuss Rawls's political constructivism in two forthcoming articles: "Rawls as Reader of Kant's Political and Moral Philosophy," in *John Rawls*, edited by Joan Vergés Gifra and Hugo Seleme (Barcelona and Buenos Aires: Katz Editores), and "Rawlsians and Other Kantians," in *The Kantian Mind*, edited by Sorin Baiasu and Mark Timmons (Routledge). In the latter article I discuss how Rawls interprets Kant's conception of objectivity and adapts it for purposes of his own political philosophy.

³ Kleingeld and Willaschek apparently belong to this group of writers, as I show below, in Part Eight. Wood also apparently belongs to this group; see, for example, KE, p. 46, 51–54.

⁴ This article is published in *Philosophers' Imprint* 19(6), February 2019.

⁵ Hereinafter, by "K&W" I refer either to the co-authored article or to the authors of the article (as its co-authors).

⁶ Here I follow the practice of K&W in capitalizing the phrase "Moral Law" in order to distinguish the law expressed by the supreme principle of morality from subordinate moral laws. Similarly, I capitalize the phrase "Categorical Imperative," which names the supreme principle of morality, in order to distinguish it from subordinate categorical imperatives.

According to K&W, although the scholars whose writings they survey take opposing sides in the debate about the source of the normativity of the Moral Law, they all share a crucial premise, namely, that the Moral Law must be either self-legislated or else legislated by another (i.e., imposed on us by some authority or source of normativity distinct from our own will, such as God, nature, or tradition). K&W reject this premise, asserting that the Moral Law is neither self-legislated nor legislated by another, but instead is a fundamental *a priori* principle of practical reason. Further, K&W declare that their arguments bring to light what they call a third option, namely, that “Kant defends an alternative to both realism and constructivism.” (K&W 1B, 4B-5 A, 14B, 16 A-B).⁷

K&W claim that all of the scholars they discuss use “the standard reading” of passages of text in which Kant writes about autonomy. I disagree with K&W’s attribution of “the standard reading” to Rawls;⁸ for this reason, I will continue to use quotation marks around this phrase. According to K&W, scholars who accept “the standard reading” interpret those textual passages as claiming that autonomy is self-legislation of the Moral Law; that “we give the Moral Law to ourselves,” or, “[m]ore precisely, *our will* gives the Moral Law *to itself*,” and that Kant’s conception of autonomy concerns “the origin of the authority of the principle of morality (the Moral Law).” K&W claim that interpreters who read Kant in this way share the “assumption that there must be *something* in which the Moral Law is *grounded*.” (K&W 1B, 14B, 16 A).

As K&W interpret Kant, he defends “the view that the Moral Law is not grounded in anything, since it is a basic *a priori* principle, as basic as pure practical reason itself.” With regard to autonomy, K&W declare that “the point of Kant’s thesis of the autonomy of the will is not to ground the authority of the Moral Law, but [instead] to indicate that all substantive moral laws are based in our own will or practical reason.” (K&W 2 A, 5 A, 13 A, 14B, 15B, 16 A-B). According to K&W, “in Kant’s view the idea of moral autonomy concerns only substantive moral laws (in the plural), such as the law that one ought not to lie (G 4:389)⁹ or ‘the law to promote the happiness of others’[.]” (K&W 2A). So, they contend,

when Kant writes that the principle of morality is the “*principle of autonomy*”, this phrase indicates not that the Moral Law itself is self-legislated but rather that the highest moral principle “*commands*” autonomy (G 4:440). Moreover

⁷ Due to the unusual page numbering practice of the journal, *Philosophers’ Imprint*, I use “A” and “B” to refer respectively to the page on the left or the page on the right side of each of the cited pages of K&W’s article.

⁸ K&W attribute “the standard reading” to Rawls early in their article (K&W 1B, 5 A), but much later (K&W 15B) they write that Rawls’s constructivist interpretation “is silent on the status of the Categorical Imperative and the Moral Law itself, since it neither claims nor denies that the Moral Law is self-legislated.” In fact, it is not silent, as I show below.

⁹ Herein, “G 4:xxx” (where each “x” represents some numeral) refers to a page of Kant’s *Groundwork of the Metaphysics of Morals*. When quoting Kant, in general (except in certain cases pointed out below) I use the standard English-language edition: the *Cambridge Edition of the Works of Immanuel Kant*, edited by Paul Guyer and Allen Wood, specifically the volume *Practical Philosophy* (Cambridge University Press, 1996), which presents Mary J. Gregor’s translations. When citing texts (by authors other than Kant) within the main text instead of in footnotes, I follow the practice of consolidating the page references at the end of a paragraph (if it contains references to only one text), except when it seems important to cite each paraphrased or quoted sentence separately.

(v), Kant's claim here is that we should act "*as if*" we were giving universal laws through our maxims. He never writes that *we*, human beings as such, *actually* give substantive moral laws; rather, these laws have their source in practical reason. (K&W 2A, italics in original text).

K&W's account of Kant's idea of autonomy, which they present as an alternative to both realism and constructivism, is extremely brief; it consists of little more than the above claims, and they offer little argumentation to support them. K&W support their account mainly by disputing other scholars' readings of short passages of Kant's texts, and they dispute them without explicitly discussing and adequately clarifying the meanings of centrally relevant terms (including "legislate," "self-legislate," "give law," "obligation," "the will," and "practical reason"), as used variously by themselves, by Kant, and by the other scholars. Therefore, their own interpretation is unclear and they do not demonstrate its superiority.

I endorse K&W's aim of correcting misrepresentations of Kant's moral philosophy (which are rife); I applaud their success in providing plausible alternative readings of the relevant passages of Kant's texts; and I appreciate their having done this scholarly work.¹⁰

Further, I agree with K&W regarding the area of intellectual territory where a satisfactory interpretation of Kant's idea of autonomy is most likely to be developed or discovered. Below I show that Rawls's interpretation is neither "constructivist," as K&W evidently construe this term, nor "realist," as they evidently construe this term, but instead is of the "third" type toward which they gesture. Rawls occupies the area of intellectual territory to which K&W apparently attempt to lay claim.

K&W dismiss Rawls's interpretation despite saying almost nothing about it and offering surprisingly little textual evidence to support what they do say. They cite only one of Rawls's articles, "Kantian Constructivism in Moral Theory" (published in 1980), which includes less than two pages about autonomy. They neglect a relevant article, "Themes in Kant's Moral Philosophy" (published in 1989), as well as the scholarly, complex, and illuminating account of autonomy that Rawls develops over the course of ten chapters of his book, *Lectures on the History of Moral Philosophy*, published in the year 2000¹¹ (hereinafter, "LHMP"). Rawls's interpretation draws from all of Kant's major works of moral philosophy, as well as from other works including *The Critique of Pure Reason*. My account of Rawls's interpretation is necessarily incomplete and selective, but I hope it suffices for present purposes.

¹⁰ K&W undertake to discredit "the standard reading" partly by pointing out that Kant does not explicitly state this or that, which one or another Kant scholar infers. In light of this, I would like to point out that, if I have inferred correctly one of K&W's importantly relevant yet tacit thoughts (i.e., that they aim to offer only as much of their own interpretation as they think suffices for their article's main purposes), then I have provided some support for the view that sometimes a reader infers correctly something not explicitly stated in a text. Beyond making this point, I do not engage here with K&W's criticisms of other scholars' readings of particular excerpts from Kant's texts. I share K&W's concern about interpretations creating the misleading impression that Kant's moral philosophy is paradoxical. An especially influential interpretation of this type, which K&W do not mention, is: Johnson, Robert; Cureton, Adam, "Kant's Moral Philosophy", *The Stanford Encyclopedia of Philosophy (Spring 2021 Edition)*, Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/spr2021/entries/kant-moral/>>.

¹¹ Harvard University Press.

K&W apparently (judging from their criticisms of rejected interpretations) hold the view that an adequate interpretation of Kant's idea of autonomy must at least fulfill the following criteria or desiderata. (1) It must ascribe to Kant neither (a) any form of voluntarism, nor (b) any problematic form of moral realism. (2) It must provide a satisfactory account of Kant's conception of moral objectivity; more specifically, it must explain Kant's view of the relation between correct moral judgments and the normativity of the supreme principle of practical reason (i.e., the universal and unconditional normativity of the Categorical Imperative expressing the Moral Law). K&W do not cite any interpretation that they regard as fulfilling their desiderata. In light of this, if an interpretation appears to fulfill them at least as well as K&W's own interpretation (or any developed by either one of these co-authors), then there are reasons why both Kleingeld and Willaschek should pay duly respectful attention to it.

Kant scholars should pay duly respectful attention to Rawls's interpretation of Kant's idea of autonomy because it fulfills K&W's desiderata to a significant degree, as I show below. Important related questions, including (a) whether it fulfills them adequately, (b) what desiderata (including those above, as there formulated, or excluding or reformulating one or more parts of them) must be fulfilled by an adequate interpretation, and (c) whether Rawls's interpretation fulfills the (other or further) desiderata are questions that go beyond the scope of this article. Also beyond the scope of this article is whether Kant's moral philosophy, as Rawls interprets it, provides a fully satisfactory conception of moral objectivity.

I argue below that K&W's main claims (understood as I have presented them, above), are false. It is not the case that all scholars (among those cited by K&W) who interpret Kant as a moral constructivist ascribe to him the paradox of self-legislation. Rawls is a counterexample. More specifically, Rawls does not ascribe to Kant the claim that the source of the normativity of the Moral Law is the law-giving or self-legislation of the autonomous person (as this claim is evidently interpreted by K&W). Further, it is not the case that all scholars (among those cited by K&W) who take opposing sides in the scholarly debate about the source of the normativity of the Moral Law share the premise that the Moral Law must be either self-legislated or else legislated by another; again, Rawls is a counterexample.

By showing that Kant's moral philosophy, as Rawls interprets it, is constructivist (as this term is used by Rawls, not by K&W), and by displaying some of its objective aspects, I offer some support for the view that a moral philosophy can be both constructivist and objective. In recent years, some writers have argued that a metaethical realist, moral realist, ethical realist, or moral objectivist philosopher can endorse a moral constructivist philosophy of some form.¹² Here I do not engage in any debate about how to define these metaethical terms. In view of the objectivist

¹² "Can Kantian Constructivism Avoid Realist Commitments?" by Michael Lyons, in *Reason, Normativity and Law: New Essays in Kantian Philosophy*, edited by Alice Pinheiro Walla and Mehmet Ruhi Demiray (Cardiff: University of Wales Press, 2020). See also, "Moral Objectivity: A Kantian Illusion?" by Carla Bagnoli, in *Journal of Value Inquiry* (2015) 49:31–45.

aspects of Rawls's interpretation, I regard Kant's moral philosophy (as interpreted by Rawls) as a form of objective moral constructivism.

Determining whether Kant's moral philosophy is a form of objective moral constructivism requires, among other things, correctly understanding Rawls's interpretation of Kant's moral philosophy. Below I argue that K&W profoundly misunderstand and misrepresent Rawls's interpretation. Further, determining whether Kant's moral philosophy is a form of metaethical realism, as argued by Allen W. Wood, requires, among other things, correctly understanding Wood's interpretation of Kant's moral philosophy. Below I also argue that K&W profoundly misunderstand and misrepresent Wood's interpretation.

Next, Parts Two and Three prepare the way for the account of Rawls's interpretation of Kant's moral philosophy presented in Parts Four through Six. Part Two discusses selected passages of Kant's *Groundwork of the Metaphysics of Morals* ("Groundwork") and argues that in order to understand the idea of autonomy, it is necessary to understand Kant's complex conception of the will. Part Three presents an overview of Kant's conception of the will. Parts Four through Six explain how, according to Rawls, Kant's conception of the will structures his constructivist moral philosophy, including its idea of autonomy. Part Seven argues that unlike Rawls, K&W apparently conflate questions about the Moral Law with importantly different questions about moral obligation that Kant distinguishes. Part Eight examines and rebuts K&W's criticisms of Rawls's interpretation. Part Nine argues that K&W misrepresent Allen Wood's account of autonomy. Part Ten concludes.

2 Kant on Autonomy and Morality

Here below I quote certain relevant passages from Kant's *Groundwork of the Metaphysics of Morals*, 4:439–441, changing the order of some sentences (without, I believe, distorting any meanings) in order to highlight certain elements of Kant's view.

Autonomy of the will is the property [*Beschaffenheit*] of the will by which it is a law to itself (independently of any property of the objects of volition). The principle of autonomy is, therefore: to choose only in such a way that the maxims of your choice are also included [*mit begriffen seien*] as universal law in the same volition. (G 4:440).¹³

The following footnote is attached by the translator, Mary J. Gregor, to the word "choice" in the second sentence quoted above:

¹³ Here, quoting G 4:440, I use Gregor's translation, interpolating the German phrases because she provided them in footnotes. Compare the translation of the same sentences by Jens Timmermann in the revised edition (Cambridge University Press, 2012): "Autonomy of the will is the characteristic of the will by which it is a law to itself (independently of any characteristic of the objects of willing). The principle of autonomy is thus: not to choose in any other way than that the maxims of one's choice are also comprised as universal law in the same willing." Timmermann does not attach any footnotes to these sentences.

zu wählen also so, dass die Maximen seiner Wahl. Kant has apparently not yet drawn the distinction between *Wille* (“the will”) and “*Willkür*” (“choice” or “the power of choice”) so prominent in the *Metaphysics of Morals*.

However, the distinction between *Wille* and *Willkür* evidently plays an important role in the *Groundwork*, since one must employ it in order to make sense of what Kant says in that work.¹⁴ In particular, that distinction is implicit in Kant’s statement that the autonomous will is a law to itself. My reasons for thinking so derive partly from the accounts of Kant’s conception of the will offered by Rawls and by Wood (discussed below, in Part Three). Rawls’s interpretation of the idea of autonomy uses Kant’s distinction between *Wille* and *Willkür*.

The pair of sentences quoted above (from G 4:440) begin a paragraph. In the rest of the same paragraph (which is the last one of G 4:440), Kant explains that “mere analysis of the concepts of morality” can prove that the principle of autonomy “is the sole principle of moral science,”¹⁵ but cannot prove that “this practical rule is an imperative, that is, that the will of every rational being is necessarily bound to it as a condition.”

Immediately prior to the above-quoted sentence that begins, “Autonomy of the will is the property [characteristic] of the will...,” Kant distinguishes between “our own will” and a “will possible for us in idea”:

Our own will insofar as it would act only under the condition of a possible giving of universal law through its maxims -- this will possible for us in idea -- is the proper object of respect; and the dignity of humanity consists just in this capacity to give universal law, though with the condition of also being subject to this very lawgiving. (G 4:440).¹⁶

The sentence quoted above distinguishes between one’s own will and the idea of a possible will, more specifically, the idea of one’s own will as it would be if it were ideally good, i.e., if one were perfectly virtuous.¹⁷ This distinction plays an important role in Rawls’s interpretation of Kant’s moral philosophy.

In the first paragraph of G 4:441, Kant contrasts autonomy to heteronomy, as follows:

If the will seeks the law that is to determine it *anywhere else* than in the fitness of its maxims for its own giving of universal law -- consequently if, in going beyond itself,

¹⁴ For relevant discussion, see *Immanuel Kant’s Groundwork for the Metaphysics of Morals: A Commentary*, by Dieter Schönecker and Allen W. Wood (Harvard University Press, 2015), especially pp. 98–103.

¹⁵ This is Timmermann’s translation. Compare Gregor’s: “is the sole principle of morals.”

¹⁶ This Gregor’s translation. Compare Timmermann’s: “Our own will, in so far as it would act only under the condition of a possible universal legislation through its maxims -- this will possible for us in the idea -- is the actual object of respect, and the dignity of humanity consists in just this capability, to be universally legislating, if with the proviso of also being itself subject to precisely this legislation.”

¹⁷ See Kant’s discussions of virtue in G 4:407, 426, 436, 442. Notice also “our ideal of moral perfection” (G 4:408). Notice the important implications for Kant’s views about humanity and dignity; I discuss them in a manuscript in progress.

it seeks this law in a property of any of its objects -- *heteronomy* always results. The will in that case does not give itself the law; instead the object, by means of its relation to the will, gives the law to it. This relation, whether it rests upon inclination or upon representations of reason,¹⁸ lets only hypothetical imperatives become possible: I ought to do something *because I will something else*. (G 4:441).

The above-quoted sentences about heteronomy evidently refer to the will of an actual individual person (any person) and distinguish between the motivational structure of a person whose will is not determined by the moral law, and that of a person whose will is determined by the moral law. This distinction between motivational structures is central to Rawls's interpretation of Kant's philosophy as a form of moral constructivism, as shown below.

According to Kant, the moral quality (worthiness) of a person's actions (taken both singly and collectively) depends on that person's motivations; morally imperfect persons (such as all human beings are, to some degree) can and should use the idea of the character of a morally ideal or virtuous agent (meaning a human agent whose will would be as perfectly and unconditionally good as humanly possible) as a guide to correct moral reasoning; further, the idea of the character of a morally ideal agent includes a conception of such an agent's motivational structure.

The essence of things is not changed by their external relations; and that which, without taking account of such relations, alone constitutes the worth of a human being is that in terms of which he must also be appraised by whoever does it, even by the supreme being. *Morality* is thus the relation of actions to the autonomy of the will, that is, to a possible giving of universal law through its maxims. (G 4:439, emphasis in original text).

In order to understand Kant's idea of autonomy, it is evidently necessary to understand his complex conception of the will, which I begin discussing next, in Part Three.

3 Kant's Conception of the Will

The second half of the chapter on autonomy in Wood's book, *Kantian Ethics*,¹⁹ examines Kant's conception of the will. Below, Part (3.1) presents elements of Wood's account of the will, which illuminates Rawls's view, presented in Part (3.2).

3.1 Wood on Kant on the Will

Wood writes:

¹⁸ Here Kant presumably does not refer to ideas of reason such as the moral law and the realm of ends, but instead to rational-intuitionistic conceptions of an objective order of values, including that of Leibniz, as discussed by Rawls, LHMP, p. 236.

¹⁹ Cambridge University Press, 2008.

[W]e are not to think of the [moral] law as legislated by our fallible, corrupt actual wills but by the pure concept of our rational will, by our will as it would be if its volitions always accorded with the rational principles it recognizes as objectively binding. (KE 111).

Here Wood, following Kant, is using three interrelated yet significantly different distinctions: between (a) the pure concept of a rational will, and the will of an actual human being; (b) an actual person's imperfectly virtuous will, and the same person's will as it would be if the person were perfectly virtuous; and (c) Kant's concepts, *Wille* and *Willkür*.²⁰

Wood notes that the traditional Latin translation of "Willkür" is *arbitrium* (conventionally translated into English as "choice"),²¹ but that of "Wille" is *voluntas* (a narrow sense of "will"). Distinguishing the two concepts, Kant writes that choice is directed to actions and their maxims, while *Wille* is not directed to actions but "immediately to giving laws for the maxims of actions (and is, therefore, practical reason itself)."²² Kant also uses "Wille" in a broad sense, encompassing both will in the narrow sense and choice, when speaking of the will as the faculty (*Vermögen*) of practical reason. (KE 16, 121).

Wille in the narrow sense does not choose but "presents to choice the reason or ground for choosing this over that," while *Willkür* chooses subjective principles of action. (KE 121). *Willkür* is subject to the laws given by *Wille*, the highest norm of which is the moral law. Wood explains:

[When an individual human person's] "choice" is tempted not to obey the [moral] law but is inwardly constrained to do so, it is constrained by (its own) will, and that is what makes it *autonomous* in obeying the moral law. (KE 121).

As Wood points out, the second part of the *Groundwork* is based on Kant's "philosophical account of the faculty of will or practical reason" and "the norms (technical, pragmatic, and moral) that are constitutive of it." Wood explains that a faculty is, roughly, "the way that a living being achieves something through processes or actions that are normatively conceived and normatively guided;" that human faculties include understanding, imagination, judgment, and theoretical reason; and that Kant regards the will as "a faculty with a nature, constituted by its function and by the norms governing its proper exercise." The function of the faculty of reason is to regulate all the other faculties, and "the sole source of all normativity" is "the supreme principle of reason." Although the source of moral norms is "in the nature" or "the essence" of practical reason, with which Kant identifies the will, "voluntarism is a fundamentally erroneous representation of where rational norms come from" and "all voluntarist interpretations of Kantian autonomy ... get the doctrine ... wrong." (KE 113–116, 118).

²⁰ Prior to Kant, the scholastic tradition of philosophy used this distinction, as Wood points out (KE 121).

²¹ Rawls uses the phrase "elective will" -- see below.

²² Kant, *Metaphysics of Morals* 6:226, quoted by Wood, KE 121.

On this last point (about voluntarism), K&W's interpretation agrees with Wood's. So does Rawls's interpretation, as I show below.

3.2 Rawls on Kant on the Will

According to Rawls, Kant undertakes "to study the principles of a pure will and to set out how persons with a fully effective pure will would act, and to ascertain what the structure of their desires as governed by the principles of practical reason would be." As conceived by Kant, a person with such a will, although affected by needs and natural inclinations, as any human being must be, would not always follow them (nor act on any other object-dependent²³ desires); whenever necessary in order to respect the requirements of the moral law, such a person would act from the principles of a pure will. Although Kant does not deny that all desires have psychological strength, "he insists on the distinction between the strength of conception-dependent and principle-dependent desires and the regulative priority -- [Bishop Joseph] Butler would say the authority -- that their corresponding principles have for ideal agents." The psychological strength of an ideal agent's conception-dependent and principle-dependent desires "exactly parallels the regulative priority of the corresponding principles of practical reason." (LHMP 151–152).

According to Kant, explains Rawls, everyone normally has "a power to elect which of our many (often contending) object-dependent desires we are to act from, or to reject them all entirely, as moved by principle-dependent and conception-dependent desires." Kant "always takes for granted" that "all persons (barring the mentally retarded and the insane) acknowledge the supreme principle of practical reason as authoritative for their will." We have "the capacity to stand above and to assess our object-dependent desires," and "[t]his gives us an elective power to determine from which of those desires, if any, we shall act." (LHMP 149, 151–152).

As Rawls points out, Kant distinguishes between acting from interest and taking an interest in an action. Where reason merely supplies a practical rule for meeting a need of inclination, what interests the agent is the object of the action, insofar as it is pleasant, meets a need, or satisfies an inclination; in such a case, the agent acts from interest; but where the will depends on the principles of reason itself, as in the case of an action done for the sake of duty, the agent takes a practical interest in the action itself. One's action has moral worth or morality,²⁴ as distinct from mere legality, when one takes a practical interest in the action itself "as correctly answering, or so we think, the principles of practical reason." (LHMP 177).

The question of how, according to Kant, one ought to employ one's elective power leads Rawls to develop his account of the categorical imperative procedure (or, as

²³ Rawls contrasts object-dependent desires to "principle-dependent" and "conception-dependent" desires. He explains "object-dependent" desires as follows: "Think of all the desires that affect us, and that contend within our person, as object-dependent desires. These are like Kant's inclinations and impulses generated in us by everything from our bodily wants and needs to social processes of learning and education." (LHMP 150–151).

²⁴ Kant writes that morality is the relation of actions to the autonomy of the will (G 4:439).

he calls it, the “CI-procedure”). Rawls takes it to “represent in procedural form all the requirements of practical reason (both pure and empirical) as those requirements apply to our maxims.” This procedure “specifies the content of the moral law as it applies to us as reasonable and rational²⁵ persons in the natural world, endowed with conscience and moral sensibility, and affected by, but not determined by, our natural desires and inclinations.” Rawls uses his account of the CI-procedure to elucidate Kant’s conception of “the structure of motives in a person with a firm good will,” his view that the moral worth or morality of actions depends on their principles of volition, and his distinction between legality and morality. (LHMP 164–165, 177).

Rawls’s account of the CI-procedure fills nine chapters of LHMP. Below I summarize this account only to the extent necessary (and, I hope, sufficient) for rebutting K&W’s inaccurate characterization of his view.

4 The CI-Procedure, Pure Practical Reason, and Autonomy

According to Rawls, applying the Categorical Imperative to ourselves “as human beings situated in our social world” requires working through the CI-procedure. In offering his account of it, Rawls aims to elucidate the themes and features that distinguish Kant’s view and, by giving them a more specific sense, to “bring to life and to make intelligible Kant’s characteristic and deeper ideas,” such as the unity of reason and acting under the idea of freedom. This is the “main value for us” of studying the CI-procedure, since it is not adequate, Rawls says, for the purpose of “generating the content -- the first principles²⁶ along with the essential rights, duties, permissions, and the rest -- of a reasonable moral doctrine” although “it is surely highly instructive as one of the more, if not the most, illuminating formulations of the requirement to express our reasons universally when assuming a moral point of view (an idea that goes back at least to Leibniz and Clarke).” (LHMP 162–163).

Rawls interprets Kant as assuming that people normally acquire a conscience and a capacity for moral judgment, and that their moral sensibility makes them pause to reflect when uncertain about moral questions confronting them due to their own and others’ needs and suffering, their own inclinations and impulses, and so on. At such times one critically examines one’s own aims and motivations, what one is considering doing and what reasons appear to justify doing so, i.e., one assesses the maxims or subjective principles implicit in one’s actions.

Rawls explains:

By representing the requirements of practical reason, the CI-procedure articulates a mode of reflection that Kant thinks can help us to gain clarity about such matters, given our moral sensibility and capacity for judgment as devel-

²⁵ Rawls uses the phrase “reasonable and rational” to express the meaning of Kant’s term *vernünfftig*, and also to mark Kant’s distinction between pure and empirical practical reason: the term “reasonable” corresponds to pure practical reason, which is expressed in the categorical imperative, and the term “rational” corresponds to empirical practical reason. (LHMP 164).

²⁶ Notice that “first principles” does not refer to the Moral Law.

oped, as it must be, in growing up and living in society. [paragraph break] It is a serious misconception to think of the CI-procedure as an algorithm intended to yield, more or less mechanically, a correct judgment. There is no such algorithm, and Kant knows this. (LHMP 166).

Rawls interprets the CI-procedure as a means of characterizing the reasoning and judgment of ideal human agents, who are “lucid and sincere,” in the sense that they know, or can formulate and state when appropriate, the reasons from which they act, and who use both the categorical imperative and the hypothetical imperative, implicitly, intuitively, and correctly, in their moral thought and judgment. So understanding Kant, Rawls takes the CI-procedure to “represent in procedural form all the requirements of practical reason (both pure and empirical) as those requirements apply to our maxims.” (LHMP 164–166, 169, 178).

A maxim, Rawls explains, is an expression of “the agent’s personal intention,” in the form of a “particular hypothetical imperative.” Our maxims are subjective principles that we have arrived at in view of what we regard as the relevant features of our circumstances, or that we might arrive at, given the normal circumstances of human life. Rawls takes Kant to assume that any maxim to which an agent applies the requirements of practical reason via the CI-procedure is rational “from the agent’s point of view: that is, the maxim is rational given the agent’s situation and the available alternatives, together with the agent’s desires, abilities, and beliefs (taken to be rational in the circumstances).” (LHMP 167–168).

Rawls interprets the CI-procedure as having four steps, which he describes as follows.²⁷ The first step is to articulate the agent’s maxim in the following standard form:

- (1) I am to do X in circumstances C in order to bring about Y unless Z. (Here X is an action and Y is an end, a state of affairs).

The second step is to generalize the maxim; “the result is what we may call a universal precept (not Kant’s terminology) that applies to everyone”:

- (2) Everyone is to do X in circumstances C in order to bring about Y unless Z.

The third step is to transform the universal precept into an “as-if law of nature”:

- (3) Everyone always does X in circumstances C in order to bring about Y, as if by a law of nature (as if such a law was implanted in us by natural instinct).²⁸

²⁷ The four steps are presented on pp. 168–169 of LHMP.

²⁸ Here Rawls cites G 4:422–423.

The fourth step is complicated; Rawls presents “the intuitive idea” as follows:

- (4) We are to adjoin the as-if law of nature at step (3) to the existing laws of nature (as these are understood by us) and then to think through as best we can what the order of nature would be once the effects of the newly adjoined law of nature have had sufficient time to work themselves out.

Here one assumes that the new order of nature has “a settled equilibrium state the relevant features of which we are able to figure out.” Calling this new order of nature an “adjusted social world,” Rawls explains that we are to associate it with the maxim at step (1), thus imputing to the agent (oneself) “a legislative intention, an intention as it were to legislate such a world.” (LHMP, p. 169).

Kant’s categorical imperative can now be stated as follows: We are permitted to act from our rational and sincere maxim at step (1) only if two conditions are satisfied:

First, we must be able to intend, as sincere, reasonable, and rational agents, to act from that maxim when we regard ourselves as a member of the adjusted social world associated with it, and thus as acting within that world and subject to its conditions; and

Second, we must be able to will this adjusted social world itself and affirm it should we belong to it.” (LHMP 169).²⁹

This rendering of the CI-procedure draws on the law of nature formulation of the categorical imperative (“FLN”): “Act as if the maxim of your action were to become through your will a universal law of nature.” (G 4:421). Noting that Kant offers three different (families of) formulations of the categorical imperative, Rawls addresses the question of how the formula of humanity as an end in itself (“FH”)³⁰ and the formula of autonomy with its variant, the formula of the realm of ends (“FA,” “FRE”), are related to the CI-procedure. He conjectures³¹ that they “depend on the CI-procedure and its content -- the maxims it accepts -- as already laid out,”

²⁹ It is interesting to compare this interpretation to those offered by Kleingeld, in her article, “Contradiction and Kant’s Formula of Universal Law” (*Kant-Studien*, 2017) and Timmermann, in his article, “A Tale of Two Conflicts: On Pauline Kleingeld’s New Reading of the Formula of Universal Law” (*Kant-Studien*, 2018).

³⁰ The following three “formulations of the moral law” are quoted from Wood’s reference list of formulas and propositions in KET, p. xx. The formula of humanity as an end in itself (FH): “*So act that you use humanity, whether in your own person or that of another, always at the same time as an end, never merely as a means*” (G 4:429; cf. 4:436; italics in Wood’s text). The formula of autonomy (FA): “...the idea of the will of every rational being as a will giving universal law” (G 4:431; cf. 4:432); or “Choose only in such a way that the maxims of your choice are also included as universal law in the same volition” (G 4:439; cf. 4:432, 434, 438). The formula of the realm of ends (FRE): “Act in accordance with the maxims of a universally legislative member of a merely possible realm of ends” (G 4:439; cf. 4:432, 437, 438).

³¹ Acknowledging that his conjecture departs from Kant’s text in a certain respect, Rawls says that he does not think it distorts Kant’s main point. (LHMP, p. 183).

and that “[t]he point of the other formulations” concerns the relevance of different points of view: using FUL/FLN, “we look at a moral situation from the agent’s point of view;” using FH, we view ourselves and other persons as passive or “as affected by our proposed action”; using FA/FRE, “we come back again to the agent’s point of view, but this time not as someone subject to moral requirements, but as someone who is, as it were, legislating universal law” or “as it were, legislating moral requirements.” (LHMP 182–183, 200).

If his conjecture is correct, Rawls says, then what it means to treat humanity as an end-in-itself “is explicable in terms of the maxims accepted by the CI-procedure, once we look at that procedure from the point of view of ourselves and others as affected by the proposed action, and so as passive.” Moreover, if we were to adhere to the CI-procedure “with a full grasp of its meaning” as expressed in all three formulations, then we could regard ourselves as “making universal law for a possible realm of ends.” (LHMP 187–188, 200).

The choice to check whether one can incorporate an inclination into a permissible maxim is “an exercise of elective will.” A person who lacked moral sensibility and did not take any practical interest in the moral law would not be led to check, via the CI-procedure, the permissibility of a tempting action. Insofar as one has a good will, the practical interest one takes in the moral law itself has “an effective regulative priority.” The action of “a person with a fully good will -- someone whose character is marked by the primary virtues of wisdom, justice, and benevolence and supported by the secondary virtues -- often involves” more than one type of interest, including a practical interest taken in the action itself; this is the case whenever persons of good will act “within their rights” in order to secure interests “but only after they have confirmed that their interests are compatible with the rights of others.” The various interests that move one’s actions are “arranged in a certain structure,” and the nature of this motivational structure “is best seen in how we work through the categorical imperative procedure.” (LHMP 178).

Rawls’s account of the CI-procedure is not complete, as he himself points out, saying that it omits “many difficult points of interpretation;” however, he thinks “it does not greatly matter,” for the purpose of understanding Kant’s argument that there is pure practical reason, “what specific formulation [i.e., account] of the CI-procedure we adopt, provided that it meets [four] essential conditions,” which include a requirement to exhibit the Categorical Imperative as a principle of autonomy showing us that we are free (LHMP 162–163, 254).

These four conditions are:

- (1) *The content condition*: “the categorical imperative procedure must not be merely formal but have sufficient structure to specify requirements on moral deliberation so that suitably many maxims are shown to be fit or unfit to be made universal law,” since otherwise “the categorical imperative as it applies to us would be empty and without content, and so also the moral law.” (LHMP 163).
- (2) *The freedom condition*: “the categorical imperative procedure must exhibit the categorical imperative, and similarly the moral law, as a principle of autonomy, so that from our consciousness of this law as supremely authoritative and regula-

- tive for us (as it applies to us via the categorical imperative and its procedure), we can recognize that we are free.” (LHMP 163).
- (3) *The fact of reason condition*: “our consciousness of the moral law as supremely authoritative for us must be found in our everyday moral thought, feeling, and judgment; and the moral law must be at least implicitly recognized as such by ordinary human reason.” (LHMP 255).
 - (4) *The motivation condition*: “our consciousness of the moral law as supremely authoritative for us must be so deeply rooted in our person as reasonable and rational that this law by itself, when fully known and understood, can be a sufficient motive for us to act from it, whatever our natural desires.” (LHMP 255).

As Rawls explains, Kant conceives the moral law as a law of freedom; he believes that thinking of it as exemplified in righteous acts “done with a steadfast mind in a complete disregard of any advantage in this or another world...uplifts [one’s] soul” and makes one aware of the dignity of one’s nature as a free person, and that “the more clearly the moral law is presented to us as an idea of reason, and the more clearly we understand its origins in our person as free, the more forcefully it arouses our moral sensibility” and moves us to “act from it.” Kant’s reason for offering several different formulations of the categorical imperative is that no single formulation is as effective as all of them together in motivating action; moreover, there is a “natural progression” from one formulation of the categorical imperative to the next, in that FH depends on FUL/FLN and FA/FRE depends on the preceding formulations and unites them in the idea of autonomy. (LHMP 201–202, 204).

Correctly understanding how Rawls sees the CI-procedure is necessary for understanding how he interprets Kant’s idea of autonomy. Next, in Part Five, I clarify the constructivist aspect of Rawls’s interpretation of Kant’s moral philosophy.

5 The Moral Law, Categorical Imperatives, and Autonomy

Rawls distinguishes the Moral Law from both the Categorical Imperative and the CI-procedure (although he does not capitalize these words). “The moral law is an idea of reason,” which “specifies a principle” that applies to all beings with the relevant reasoning capacities. This principle is the Categorical Imperative; human beings experience it as a constraint. The CI-procedure adapts the Categorical Imperative to the normal conditions of human life “in the order of nature.” (LHMP 165, 167, 198–199, 239).

A particular categorical imperative or “practical law” is a universal precept (generalized maxim) that passes the test of the CI-procedure. Particular categorical imperatives or “precepts of morality” are “substantive principles that express the order of moral values” by “giv[ing] the content of the duties of justice and of virtue.” The Moral Law and the Categorical Imperative, “as grounded on pure practical reason,” are a priori; however, “particular categorical imperatives (particular duties of justice and duties of virtue)” are not a priori; yet they are unconditional, in contrast to hypothetical imperatives. Particular categorical imperatives give the content of the Moral Law, in the sense of specifying the duties of justice and of

virtue; “the totality of particular categorical imperatives” specifying these duties is “constructed.” (LHMP 165, 168, 237, 239, 241, 248–251).

Rawls restates Kant’s autonomy formulation of the categorical imperative as follows:

Always act so that the totality of the maxims from which you act is such that you can regard yourself as enacting through those maxims a unified scheme of public³² moral precepts the endorsing of which by all reasonable and rational persons is consistent with their humanity and would bring about (under favorable conditions) a realm of ends. (LHMP 205).

Rawls points out that “all the variants [of FA/FRE] stress the idea that we are to act in such a way that we can regard ourselves as legislating universally through our maxims,” and that some of the variants make explicit that “we are to view ourselves as members of a possible realm of ends (a moral commonwealth).” So viewing ourselves, and so acting, Rawls explains, requires us to “be sure that the maxims from which we act answer to the requirements of practical reason,” and “hence that our maxims do not subject others to purposes that do not accord with a law they can endorse as consistent with their humanity.” We are to view our maxims “as authorized by precepts that could serve as the publicly recognized moral law” of a realm of ends; thus, we are to regard ourselves “as legislating the content of the categorical imperative (as it applies to us).” (LHMP 203, 204, 205).

Rawls emphasizes that the thought of legislating for a possible realm of ends is “purely hypothetical,” and that one is to consider whether “the whole family of general precepts authorizing one’s maxims” could serve as “a moral law for a possible realm of ends.” (LHMP 204, 205). Notice that Rawls’s phrase, “a moral law,” although grammatically singular, does not here refer to the Moral Law; instead, it refers to a family or scheme of general moral precepts. Notice also that regarding ourselves as legislating the content of the categorical imperative is distinct from regarding ourselves as legislating the Categorical Imperative or the Moral Law expressed by it.

³² Regarding publicity, Rawls says: “[P]lainly Kant assumes as a law of nature that people learn from experience and remember the past; hence once it becomes, as it were, a law of nature that everyone tries to make a false promise (in certain circumstances), the existence of the law becomes public knowledge. Everyone knows of it, and knows that others know of it, and so on. We need not suppose that all laws of nature are public knowledge; obviously they are not. But as a way of interpreting the requirements of the CI-procedure in terms of the law of nature formulation, it is not inappropriate to assume the public recognition of the as-it-were laws of nature generated by people acting from certain maxims. [paragraph break] We make this explicit by saying that in the equilibrium state of the adjusted social world, the as-it-were laws of nature at step (3) [of the CI-procedure] are publicly recognized as laws of nature, and we are to apply the CI-procedure accordingly. Let’s refer to this public recognition of the as-it-were laws of nature issuing from maxims at step (1) [of the CI-procedure] as the publicity condition on universal moral precepts. Kant views acceptable precepts of this kind as belonging to the public moral legislation, so to speak, of a moral community.” (LHMP, p. 171). Rawls also specifies a further condition, which he explains on pp. 171–172. However, discussing it would require going beyond the scope of the present article.

According to Rawls, the “public moral law for a realm of ends” is constituted or constructed by each and all who “intelligently and conscientiously follow the principles of practical reason.” (LHMP 203). Although Rawls interprets Kant’s moral philosophy as in this sense a form of constructivism, he does not interpret it as a form of voluntarism. Nor does Rawls interpret it as a form of subjectivism or relativism. He emphasizes that efforts to specify the Moral Law are “always a matter of approximation, and always subject to error and correction,” by which he means that “since an idea of reason can never be fully realized, neither can the content of such an idea.” (LHMP 239, note 4).

I have shown that Rawls distinguishes between the Moral Law and moral laws and between the Categorical Imperative and categorical imperatives, unlike many of the Kant scholars criticized by K&W. Also I have argued that Rawls does not interpret Kant’s moral philosophy as a form of voluntarism. Next, in Part Six, I summarize Rawls’s account of Kant’s conception of moral obligation.

6 Rawls on Kant on Obligation and Autonomy

Kant’s conception of moral obligation, explains Rawls, is that we each and all are “bound only by a law that we can give to ourselves,” which means that we all are legitimately bound by, or subject to, the requirements of the categorical imperative “just because we can, *as reasonable and rational*, regard ourselves as legislating the content of the categorical imperative (as it applies to us),” and that all the moral norms to which we all are legitimately subject are “norms we can view ourselves as legislating *as reasonable and rational persons*.” (LHMP 205, 206; all italics added).³³ Notice that Rawls here writes about how to assess laws or norms to which we are purportedly subject, but does not write that we should judge whether we can legislate (nor whether we can regard ourselves as legislating) the Categorical Imperative. Assessing a law or norm requires judging whether it is consistent with the duties of justice and virtue. Determining these duties requires using the CI-procedure.

According to Rawls, Kant’s idea of autonomy requires that the form of the reasoning procedure for specifying the content of duties of justice and virtue (i.e., the CI-procedure) be determined by “our conception of ourselves as reasonable and rational persons (possessing the powers of practical reason) and by our conception of the public role of moral principles in a possible realm of ends.”³⁴ (LHMP 236–237). Rawls argues that if any philosophical theory were to determine duties differently, for example, as grounded in principles “founded on relations among objects the nature of which is not affected or determined” by those conceptions, then

³³ I have added the italics to the quoted passages in order to emphasize that here, when using the phrases “ourselves” and “our own will,” Rawls is employing the distinctions highlighted in the first paragraph of Part (3.1) above.

³⁴ Other influential contemporary writers about moral constructivism say less, if anything, about the public role of moral principles in a possible realm of ends.

that theory would describe practical reason as heteronomous;³⁵ the only alternative is the view “that there exists no moral order prior to and independent of those conceptions.” (LHMP, pp. 235–237). Notice the close relation between autonomy and moral constructivism.

Rawls writes that several textual passages, including KP 5:62,³⁶ show that “Kant supposes there to be only two possibilities: either the moral law is founded on an object given to it, in which case it depends on our susceptibility and the pleasure we anticipate from realizing that object, or the moral law as pure practical reason determines (constructs) its own object out of itself.” (LHMP, p. 235). Here Rawls argues, as regards the question of the Moral Law’s foundation, basis, or ground, that Kant assumes that there are only two possibilities, heteronomy and autonomy. I highlight this partly because K&W appear to argue that there is a third possibility.³⁷ I examine their views on this point, among others, in Part Nine.

Above I have summarized Rawls’s account of Kant’s conception of moral obligation, highlighting the relation between autonomy and constructivism. Below, in Part Seven, I show that Kant distinguishes the question of the nature of moral obligation from the question of the source of the normativity of the Moral Law. Unlike K&W, Rawls makes this distinction clear.

7 Moral Obligation and the Normativity of the Moral Law

K&W assert that Kant defends “the view that the Moral Law is not grounded in anything, since it is a basic a priori principle, as basic as pure practical reason itself.” (K&W 16 A). They offer this assertion as an answer to the question of the source of the “obligatory force” of the Moral Law. K&W appear to regard this phrase as synonymous with “binding force,” “normative force,” “normative validity,” and “normative authority,” all of which they use roughly interchangeably throughout their article. Since they use these phrases interchangeably, they appear to conflate questions that Kant distinguishes.

In Kant’s moral philosophy, questions about moral obligation (e.g., what moral obligation is and how it is possible) are distinct from, and answered separately from, questions about moral law (e.g., what a moral law is and what considerations

³⁵ Kant would reject the metaphysical perfectionism of Leibniz and other variants of rational intuitionism, as well as Hume’s psychological naturalism, as forms of heteronomy, Rawls argues. (LHMP 105–140, 235).

³⁶ Kant explains (in KP 5:62) that when “a determining ground of the faculty of desire precedes the maxim of the will,” the “end itself, the gratification that we seek, is in the latter case not a *good* but a *well-being*, not a concept of reason but an empirical concept of an object of feeling [...]”

³⁷ Rawls himself appears (on p. 235 of LHMP) to point out a third possibility; however, here he is merely arguing that Kant does not make clear why he regards rational intuitionism (e.g., Leibniz’s metaphysical perfectionism) as heteronomous despite its similarities to Kant’s own view as regards the relation between (a) the object of thought (in Leibniz’s case, this is the moral order that is fixed by the divine nature and is therefore prior to and independent of human moral persons), and (b) moral motivation.

establish that the Categorical Imperative expresses the Moral Law and is the supreme moral principle). It is unclear which, if any, of Kant's questions K&W answer on his behalf. However, adequately answering any of the apparently conflated questions requires offering a fuller answer than they offer.

Kant's desiderata for a satisfactory philosophical account of morals evidently include the following (stated roughly, since greater precision does not seem necessary for my present purpose): (a) The subject matter of the theory must not be too different from what most people using common sense would regard as morality. (b) The theory must clarify at least the most important of the commonly used moral concepts. (c) It must correctly identify the principles and values that ground sound moral judgments and decisions. (d) It must explain what moral obligation is.

As regards (d), if moral obligation were no different from any other kind of obligation, and if this were incontestably obvious to everyone, then it would not seem necessary to begin developing one's philosophical account of moral obligation by developing an account of morality, as does Kant. In the *Groundwork* Kant first undertakes to show that the Categorical Imperative is the principle underlying commonsense moral judgments about goodness of will and moral worth of actions, then offers his explanations of morality and moral obligation. "*Morality* is thus the relation of actions to the autonomy of the will, that is, to a possible giving of universal law through its maxims." (G 4:439).³⁸ On the basis of this understanding of morality, Kant explains (in the next sentence) what makes an action morally permissible or impermissible: "An action that can be consistent with the autonomy of the will is *permissible*; one that does not agree with it is *impermissible*" (G 4:439, italics in original). In the next three sentences he turns from the topic of moral classification of actions to the topic of the will, in order to explain moral obligation:

A will whose maxims necessarily harmonize with the laws of autonomy is a *holy*, absolutely good will. The dependence on the principle of autonomy of a will that is not absolutely good (moral necessitation) is *obligation*. This cannot therefore refer to a holy being.

Moral obligation is the relation between the human will and the Categorical Imperative, and it is a dependency relation in which the will is in the subordinate position. (G 4:439). Explaining this relation therefore requires discussing not only the Categorical Imperative but also the will and, more generally, moral psychology; moreover, the requisite account of moral psychology must resolve an apparent paradox concerning moral motivation³⁹ (a paradox of autonomy that Kant himself points out but K&W do not discuss). Furthermore, in Kant's view an adequate philosophical account of moral obligation must include not only an elucidation of the concept of moral obligation and an account of moral psychology, but also an explanation of

³⁸ The italics are in Kant's text. Compare Timmermann: "*Morality* is thus the relation of actions to the autonomy of the will, that is, to the possible universal legislation through its maxims."

³⁹ "And the paradox lies just in this: that the mere dignity of humanity, as rational nature, without any other end or advantage to be gained by it, and hence respect for a mere idea, is still to serve as an unrelenting prescription of the will...." (Kant, *Groundwork*, 4:439 [Timmermann; compare Gregor]).

how morally necessitated action is possible, which in turn requires addressing questions about freedom of the will.

In support of his claim that the Categorical Imperative is the supreme principle of morality, Kant not only offers arguments in the *Groundwork* analyzing what he regards as the universally accepted concept of morality and explaining the principles of reasoning and the concept of a categorical imperative, but also offers arguments, in this and other works, both to show that his own philosophical account of morality is superior to other philosophical accounts of it, and to establish the autonomy of reason. In the *Groundwork*, Kant's arguments in support of his claim that the Categorical Imperative is the supreme principle of morality are distinct from his arguments in support of his claims about the nature and possibility of moral obligation, and he answers the question of the nature of moral obligation after determining the supreme principle of morality; the answer to the former question depends on the latter.

An adequate interpretation of Kant's moral philosophy must be faithful to the distinction between the latter two questions. Rawls's interpretation is faithful to the distinction; K&W appear to conflate the questions. Next, in Part Eight, I examine and rebut K&W's criticisms of Rawls's interpretation of Kant's moral philosophy.

8 K&W Versus Rawls

Below I quote all the sentences in K&W's article that are explicitly about Rawls, as distinct from being explicitly about "constructivist readings" of "Kant's ethics," the category into which K&W place Rawls's interpretation. (K&W 15 A).

[Constructivist readings of Kant's ethics] come in different varieties. First, there is John Rawls's Kantian⁴⁰ constructivism, according to which moral obligations are the outcome of a hypothetical deliberative procedure defined by the Categorical Imperative (the "CI procedure", Rawls 1980). This is a claim not about what grounds the bindingness of the Moral Law but about how to establish particular moral obligations. According to Rawls, the CI procedure is not a mere epistemic tool by which we discover what is morally right. Rather, on his view, moral obligations are the outcome of this procedure. (Note that this kind of constructivism is silent on the status of the Categorical Imperative and the Moral Law itself, since it neither claims nor denies that the Moral Law is self-legislated.) (K&W 15B).

Here K&W claim to offer a brief account of "John Rawls's Kantian constructivism." This phrase is ambiguous between (a) Rawls's interpretation of Kant's moral philosophy and (b) Rawls's own political philosophy as presented in his famous first book, *A Theory of Justice* (1971), as well as a number of articles. I take K&W to be offering an account of Rawls's interpretation of Kant's moral philosophy. As such,

⁴⁰ Notice the term "Kantian," as used in this instance. Below, in Part Nine, I point out another instance of use of this term by K&W that is problematic because they do not clearly distinguish between what is "Kant's" and what is "Kantian."

their account is faulty, not least because it claims that Rawls's view is silent on the status of the Moral Law; this claim is false, for the reasons given above.⁴¹ Moreover, by "moral obligations" K&W appear to refer either to what Rawls refers to as duties of justice and virtue or else to the categorical imperatives specifying the duties. K&W neither define their own term, "moral obligation," nor explain how Kant understands the distinct concepts of moral obligation and duty; therefore it is not clear what fault they find with Rawls's interpretation of Kant, nor why they reject it, although it is quite clear that they do reject it together with the other interpretations that they classify as "constructivist."

In the following textual passage, K&W present their own view and contrast it to a view they reject.

If the Moral Law is an a priori principle of practical reason itself, the obligatory force of which we come to acknowledge in practical deliberation, this suffices to account for its universality, unconditionality, and non-heteronomous origin. And if its normative validity is something we come to acknowledge in practical deliberation, we do not first establish it through an act of the will, and thus its validity is not conditional on anything else. [*paragraph break*] (K&W 4B-5A).

Since K&W do not ascribe to Rawls the interpretation of Kant that they themselves endorse, and since they classify as either realist or constructivist the views of all of the other scholars whose views they survey, and since K&W appear to ascribe to all constructivist interpretations the idea that moral obligation depends on a "volitional act" on the part of the agent,⁴² it seems reasonable to infer that in the quoted passage above K&W implicitly ascribe to Rawls the view that, according to Kant, we "establish" the normative validity of the Moral Law "through an act of the will" prior to practical deliberation. Ascribing such a voluntarist view to Rawls would be incorrect, as I hope is clear by now.

K&W say the following, immediately after the paragraph break indicated at the end of the textual passage in the block quote above:

In other words, the assumption that the origin of the normative authority of the principle of morality lies either in autonomy or in heteronomy is based on the disputable presupposition that the obligatory force of the Moral Law must be *grounded* in something more fundamental (be it a value, as many realists assert, or a principle or activity related to agency, as constructivists assert). Beyond the autonomy/ heteronomy dichotomy lies a third, overlooked possibility — namely that Kant's Moral Law, as the most fundamental practical principle, does not have a deeper "ground". That is to say, what remains open is the possibility that the Moral Law is neither self-legislated nor legislated by someone (or something) else. (K&W 5A).

It is not clear why K&W think we should, or can, go "[b]eyond the autonomy/ heteronomy dichotomy" when interpreting Kant's moral philosophy; given the

⁴¹ See above, Part Five, and footnote #8.

⁴² See K&W, pp. 1B, 4 A, 15 A.

autonomy of reason,⁴³ where else are we to go? Since it is clear that K&W regard reason as the source of the Moral Law's normative authority, I infer that either K&W interpret Kant's idea of autonomy quite differently from Rawls (and differently from Kant, it would seem), or else they are unaware of the important similarities between their own interpretation of autonomy and that of Rawls.

Determining whether Kant's moral philosophy is a form of objective moral constructivism requires, among other things, correctly understanding Rawls's interpretation of Kant's moral philosophy. K&W misunderstand Rawls's interpretation, as I have shown. Further, determining whether Kant's moral philosophy is a form of metaethical realism (as argued by Wood) requires, among other things, correctly understanding Wood's interpretation of Kant's moral philosophy. Next I argue that K&W misunderstand Wood's interpretation.

9 K&W Versus Wood

Below I offer a brief overview of Wood's account of autonomy before examining K&W's criticisms of it. A much fuller discussion of Wood's account would be necessary for doing it justice. However, a brief overview suffices for my current purpose of rebutting K&W.

Here I refer to the account offered in Chap. 6, "Autonomy," in Wood's book, *Kantian Ethics* (op. cit.), which is the sole chapter of the sole text that K&W cite in order to support their interpretation of Wood's account of autonomy.⁴⁴ Part (3.1), above, draws from and cites the second half of the same chapter, "Autonomy." Below I draw from and cite mainly this chapter's first half.

9.1 Wood's Chapter, "Autonomy"

The structure of Wood's chapter on autonomy, as well as some of its rhetorical features, may mislead some readers. In the book's preface, Wood distinguishes "Kantian ethics" from Kant's own thought. He writes:

This book [*Kantian Ethics*] attempts to sketch an ethical theory based on the principles found in the writings of Immanuel Kant. It is not primarily a study of those writings but an attempt to develop out of Kant's thought the most defensible theory possible on that basis. (KE, ix).

However, the book's chapter on autonomy is primarily a study of Kant's writings; more precisely, the chapter's latter four parts do, although its first part does not, offer

⁴³ Regarding the autonomy of reason, see, e.g., Kant's *Critique of Practical Reason*: "The supersensible nature of [rational beings in general] is their existence in accordance with laws that are independent of any empirical condition and thus belong to the *autonomy* of pure reason." (5:43, italics in Kant's text).

⁴⁴ They also offer a single citation of an earlier book by Wood in order to support their claim that Wood defends a realist reading of Kant's ethics. They do not cite Wood's more recent publications, including the commentary on the *Groundwork* that he co-authored with Dieter Schönecker (op. cit.).

scholarly interpretations of Kant's texts. The chapter's first part is titled: "Tensions within the Idea of Autonomy." What Wood writes about in this first part is "the Kantian idea of autonomy." (KE, 106). There is, he says, "a serious tension in the Kantian idea of autonomy," which "threatens to pull the doctrine of autonomy apart, depending upon whether we emphasize the '*autos*' or the '*nomos*' -- the rational being's will as author or legislator of the moral law, or the law itself as objectively binding on that same will." (KE 106).

Noting that various writers later picked up Kant's term, "autonomy," and interpreted it "in widely different ways," Wood sketches "the history of the Kantian conception of autonomy from its first reception down to the present day." (KE 106). First he briefly describes interpretations stressing the '*autos*' element of autonomy (including those offered by the early German Romantics and the existentialists), and points out their voluntarist and paradoxical aspects. (KE 107). Next he briefly describes "constructivist" interpretations not only as emphasizing the '*nomos*' element by claiming universal validity, but also as voluntarist and paradoxical due to holding (or so Wood claims)⁴⁵ that our subjective acts can confer validity on principles and that "[e]very law or value we recognize must be constituted by our volitional act in legislating it." (KE 107–108).

Having characterized these "popular interpretations of the Kantian doctrine of autonomy," Wood rejects them in favor of what he regards as Kant's own view:

Kant himself does not look at autonomy in any such way. For him, [the] content of the moral law could not be subject to my whims or an expression of my individuality, because the law is given universally, by every rational will, and not only to itself but to all other rational beings as well. I cannot loose myself from the moral law, because it is not up to me to make or unmake the idea of a rational will. (KE 108).

In the second and third parts of the same chapter, Wood argues that a certain philosophical conception of law is "built into Kant's idea that the moral law is a law of autonomy;" more specifically, "for Kant the moral law is a natural law." (KE 108, 114). In the chapter's fourth and fifth (of five) parts, Wood discusses the nature of the will and interprets Kant's conception of morals as "the science of all the ends that are *established through the nature of the will* and prescribe the objective laws of the will, and according to which we direct and exert our faculties." (KE 114).⁴⁶ After discussing, in the chapter's fifth part, the faculty of the will and the distinction between *Wille* (*voluntas*) and *Willkür* (*arbitrium*),⁴⁷ Wood explains that when one is tempted not to obey the moral law but nevertheless obeys it because one's *Wille* constrains one's *Willkür*, one is constrained by one's own autonomous will, i.e., by practical reason. (KE 114, 121).

⁴⁵ Wood's characterization of constructivist interpretations does not accurately describe Rawls's interpretation; as I have shown above, Rawls distinguishes between the Moral Law and moral laws, and his interpretation of Kant's idea of the Moral Law is inconsistent with the thought that we constitute the Moral Law by a volitional act of legislation.

⁴⁶ Wood here quotes Kant's *Lectures on Anthropology* 25:438; the italics are in Wood's text.

⁴⁷ See above, Part (3.1).

9.2 K&W's Criticisms of Wood's Account of Autonomy

K&W write:

According to Allen Wood, *Kant's idea of autonomy* includes both the idea of morality as objectively binding and the idea of “the rational being's will as author or legislator of the moral law.” (K&W 2B, italics added; K&W quote KE 106).

Here K&W misrepresent Wood's view by quoting from a sentence in which he refers to *the Kantian idea of autonomy* (as distinct from Kant's own idea). Moreover, the quote is misleading also because K&W present it apart from its context in the chapter's first section, i.e., its introductory part, where Wood discusses the history of the reception of Kant's idea of autonomy (prior to offering his own account of Kant's idea). K&W quote a second time from the same page of Wood's text, again misleadingly:

[I]t is difficult to give coherent sense to Kant's alleged claim that the will or practical reason itself is somehow actively involved in generating the obligatory force of the Moral Law. The more one emphasizes the impersonal, a priori, timeless character of autonomy, the harder it is to make literal sense of self-legislation as an *act* or *activity* of the will. Allen Wood articulates a view held by many when he writes that there is a “serious tension in the idea of Kantian autonomy” because the idea that rational beings are themselves legislators of the Moral Law and the idea that the Moral Law is objectively binding pull in opposite directions. (K&W 3B).

Since K&W say almost nothing about the account of Kant's idea of autonomy that Wood offers in the non-introductory parts of the chapter from which they quote (merely stating briefly that he offers a non-literal reading of Kant's language of self-legislation), and since they also say almost nothing about Wood's interpretation of Kant's moral philosophy (merely stating briefly that he regards it as a realist view grounded in the value of humanity), those two quotes (from KE 106) together create the false impression that Wood's main claim about Kant's idea of autonomy is that it contains a serious tension (a claim with which K&W disagree) and that he says little else about it.

10 Conclusion

Many interpreters of Kant's moral philosophy err by failing to understand how and why he distinguishes between the Moral Law and moral laws, thus failing to interpret correctly various important passages of his texts. K&W are right to highlight such errors. Their co-authored article focuses the reader's attention on whether an interpreter does or does not commit such errors, and this perspective is illuminating. Another merit of K&W's article is that it helpfully clarifies passages of Kant's texts that have caused confusion about the Moral Law and moral laws. I appreciate their having done this scholarly work.

However, K&W incorrectly characterize Rawls's interpretation of Kant. As argued above, contra K&W, Rawls distinguishes between the Moral Law and moral laws. He also interprets Kant as arguing that the Moral Law is an idea of reason, and does not interpret Kant as saying that the Moral Law results from human legislation,

self-legislation, law-giving, decisionmaking, or choosing. Moreover, Rawls clearly distinguishes, as Kant does but K&W do not, between certain importantly different questions about normativity and obligation (including, “What is the source of the normativity of the Moral Law?” and “What is moral obligation?”). An adequate account of Kant’s idea of autonomy must distinguish these questions, and must also explain how Kant conceives the will and practical reason. Rawls’s account satisfies these desiderata.

As pointed out above (in Part One), K&W claim that although the scholars whose views they discuss take opposing sides in the scholarly debate about the character and source of the Moral Law’s normativity, they all share the premise that the Moral Law must be either self-legislated or legislated by another, i.e., that the character and source of its normativity must be explained in one of these two ways. I have argued that this claim is false; Rawls is a counterexample. I have also argued against K&W’s claim that all “constructivist” scholars (or all whose views they discuss) ascribe to Kant the paradox of self-legislation. Again, Rawls is a counterexample. Although I have not disputed the correctness of K&W’s claims about the interpretations offered by all of the other scholars whose views they briefly discuss and quickly dismiss, I have done so regarding Allen Wood’s interpretation of autonomy.

Wood, Rawls, and K&W clearly agree that studying Kant’s moral philosophy is worthwhile, partly because of the great importance, for all human lives and societies, of sound moral reasoning. I agree with them on this point. At stake in the debate about how to interpret Kant’s idea of autonomy (and in particular, about whether it is or is not paradoxical) is the coherence and cogency of Kant’s philosophical conception of reason, including his perspective on reason’s proper roles in guiding individual lives and social practices, as well as in structuring political, economic, and legal systems. Scholars should proceed very carefully, not only when interpreting Kant’s texts but also when interpreting writings by other Kant scholars who appreciate the importance of these scholarly debates.

It is noteworthy that misunderstanding Rawls’s claim that Kant’s moral philosophy is a form of moral constructivism entails misunderstanding Rawls’s account of Kant’s idea of autonomy, and vice versa. Moreover, if Rawls’s interpretation of Kant’s moral philosophy is correct, then misunderstanding Kant’s moral constructivism entails misunderstanding Kant’s idea of autonomy, and vice versa. So, if Rawls’s interpretation is correct, then the many other Kant scholars who have rejected it and/or paid little regard to it, including K&W and Wood, may have misunderstood, to some significant degree, not only Rawls’s work but also Kant’s moral philosophy.

Rawls’s account of Kant’s idea of autonomy fulfills to a high degree K&W’s desiderata for an adequate account. It also fulfills to a high degree my own desiderata. Kant scholars ought to pay duly respectful attention to it. Rawls’s insightful interpretation of Kant’s moral philosophy occupies the area of intellectual territory to which K&W attempt to lay claim, and if Rawls’s interpretation is correct, it supports the possibility of a defensible form of objective moral constructivism.⁴⁸

⁴⁸ I am grateful to Pauline Kleingeld, whose body of work on Kant I value highly, for our exchanges about his philosophy both in person and in writing. I thank Sorin Baiasu for the invitation to take part in the conference at Keele University in conjunction with Kleingeld’s 2018 Rousseau Lecture, and also for his helpful comments during that conference, his work on the special journal issue, and his editing suggestions. For helpful comments during that 2018 conference I thank the other participants, in particular Mehmet Ruhi Demiray, who provided excellent written comments, also Marie Newhouse, Paola Romero, Jens Timmermann, and Mark Timmons. I thank Allen W. Wood for helpful correspondence about Kant and metaethics. For helpful conversations about Kant and metaethics I thank my colleague, Christoph Hanisch, and my student, Sebastian Johnson.

Declarations

Conflict of Interest The author declares that she has no conflict of interest.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.