Philosophia (2023) 51:555–569
https://doi.org/10.1007/s11406-023-00642-z

Kant's Formula of Autonomy: Continuity or Discontinuity?

Pauline Kleingeld[1]

Received: 1 March 2023 / Revised: 1 March 2023 / Accepted: 3 March 2023 / Published online: 12 April 2023

Abstract
In two recent articles I have argued that Kant's legal and political philosophy can shed new light on his much-contested account of moral autonomy and that important changes in his political theory help to explain why in his later work the Formula of Autonomy disappears. In the present essay, I respond to comments by Sorin Baiasu and Marie Newhouse, who argue that the changes in Kant's political theory fail to explain the disappearance of the Formula of Autonomy, since in both phases Kant held that laws are given by the people's representatives. I offer additional support for my original argument by developing a more detailed account of Kant's conception of the relation between the legislating representatives and the people they are taken to represent. I argue that Kant's conception of the proper representation relation changed, and that it changed so fundamentally that the analogy at the basis of the Formula of Autonomy was no longer apt, thus providing a plausible explanation for its disappearance. I also address several related comments and propose an explanation for why Kant dropped the very idea of autonomy as a property of the will.

Keywords Analogy · Autonomy · Formula of Autonomy · Immanuel Kant · Legislation · Representation

1 Introduction

Kant's account of moral autonomy has been immensely influential, avidly contested, and interpreted in many different ways. In two recent essays, I have argued that Kant's legal and political theory can shed new light on his account of autonomy. In 'Moral Autonomy as Political Analogy: Self-Legislation in Kant's *Groundwork* and the *Fey-*

✉ Pauline Kleingeld
pauline.kleingeld@rug.nl

[1] Faculty of Philosophy, University of Groningen, Groningen, The Netherlands

erabend Lectures on Natural Law' (Kleingeld, 2018a), I start from the observation that Kant taught his course on legal and political theory during the very months in 1784 when he was writing the *Groundwork*, and I point out several parallels between his conception of just political legislation and the moral theory of the *Groundwork.* Kant conceives of the relation between the Categorical Imperative and substantive moral laws as being structurally the same as the relation between constitutional laws and positive state laws. He sees both the constitutional laws of the state and the Categorical Imperative as a priori principles of pure reason, and these a priori principles provide the normative criterion for positive state laws and substantive moral laws, respectively. Furthermore, the legislation analogy suggests that when Kant claims the will must be viewed as 'self-legislating' (GMS 4:431),[1] this should not be understood as the will's addressing substantive moral laws primarily *to itself* but rather as its legislating *by itself* and giving laws *to all*, including itself.

In 'The Principle of Autonomy in Kant's Moral Theory: Its Rise and Fall' (Kleingeld, 2018b), I observe that Kant hardly uses the term 'autonomy' and no longer refers to the 'Principle' or 'Formula' of Autonomy in his work from the mid- to late 1790s, even though he attributed great importance to this formula in the *Groundwork*. I offer a possible explanation, pointing to changes in Kant's political philosophy. At the time of the *Groundwork*, Kant described the ideal of just legislation in terms of an enlightened *ruler* giving just laws to his people without needing to consult them. In the *Metaphysics of Morals*, by contrast, he argues that *the people* should legislate. I argue that the legislation analogy underlying the Formula of Autonomy was no longer apt as a result.

I am very grateful to Sorin Baiasu and Marie Newhouse for their perceptive comments concerning these two essays (Baiasu, 2023; Newhouse, 2023). I am glad that they endorse my claim that Kant's legal and political theory underwent significant development from 1784 until the late 1790s and that the Formula of Autonomy as such disappears from Kant's later writings. Baiasu and Newhouse are not convinced, however, by my *explanation* for its disappearance. Their main objection is that although Kant changed his political theory in many ways, he retained the idea that legislation is enacted by *representatives* of the people. As a result, they claim, the changes in Kant's later legal and political theory do not invalidate the earlier analogy, and Kant may still have been committed to the Formula of Autonomy at the time of writing the *Metaphysics of Morals*, even if he did not make this explicit.

In this essay, I respond to their comments. I offer additional support for my original argument by developing a more detailed account of Kant's conception of the relation between the legislating representatives and the people they are taken to represent. I argue that Kant's conception of this relation changed, and that it changed

[1] All references to Kant's work cite volume and page number of the *Akademie-Ausgabe*. Translations are taken from the edition listed in the References (Kant 1996), though I have sometimes altered them. Abbreviations: GMS: Groundwork for the Metaphysics of Morals (*Grundlegung zur Metaphysik der Sitten*); KpV: Critique of Practical Reason (*Kritik der praktischen Vernunft*); MS: Metaphysics of Morals (*Metaphysik der Sitten*); RGV: Religion within the Boundaries of Mere Reason (*Die Religion innerhalb der Grenzen der bloßen Vernunft*); V-Mo/Mron: Mrongovius Lectures on Moral Philosophy (*Moral Mrongovius 2*); V-NR/Feyerabend: Feyerabend Lectures on Natural Law (*Naturrecht Feyerabend*); ZeF: Toward Perpetual Peace (*Zum ewigen Frieden*).

so fundamentally that the analogy at the basis of the Formula of Autonomy became inapt, thus providing a plausible explanation for its disappearance. I also address several related comments and propose an explanation for why Kant dropped the very idea of autonomy as a property of the will.

I start with a brief statement of the relevant parts of my main argument[2] (Sect. 1), and then I present the comments and objections formulated by Baiasu and Newhouse (Sect. 2). I first respond to their objection that the role of representatives in the legislative process does not change (Sect. 3). I subsequently discuss Kant's sporadic use of the term 'autonomy' in the Doctrine of Virtue, explaining why this does not refer to autonomy as a property of the will. I propose an answer to the question of why Kant may have given up the very idea of autonomy of the will (Sect. 4). I then discuss Kant's possible reference to the general will in the *Groundwork*, as well as the role of God in ethical legislation (Sect. 5). In the final section, I comment on the use of the terms 'moral' and 'ethical' (Sect. 6).

2 The Rise and Fall of the Principle of Autonomy

The 'Principle of Autonomy', Kant writes in the *Groundwork*, 'is the idea *of the will of every rational being as a will giving universal law*' [*eines … allgemein gesetzgebenden Willens*] (GMS 4:431, original emphasis), or, in an alternative formulation, the '*principle* of every human will *as a will that is universally legislating through all its maxims*' (GMS 4:432, original emphasis). He calls this the 'third formula' of the Categorical Imperative (GMS 4:432) and proceeds to reformulate it as the moral demand that one evaluate one's maxims by imagining oneself to be a legislator who gives laws to all rational beings through one's maxims. Each rational being 'must regard himself as legislating universally through all the maxims of his will, in order to evaluate [*beurteilen*] himself and his actions from this point of view' (GMS 4:433). The Formula of Autonomy articulates the demand that one 'act only so that the will can regard itself as simultaneously legislating universally through its maxim' (GMS 4:434). Kant again reformulates this later as follows: 'act as if your maxim were to serve at the same time as a universal law (of all rational beings)' (GMS 4:438).

In the *Groundwork*, Kant explains that the different formulas following the Formula of Universal Law are based on analogy (GMS 4:436, 437). And indeed, the common core of the different formulations of the 'Principle' or 'Formula' of Autonomy is a legislation analogy: for the purposes of morally evaluating your actions and the maxims that underlie them, you are to 'regard yourself as' (or 'act as if' you are) simultaneously legislating *universally*, that is, *to all*. Importantly, in imagining yourself to be giving laws *to all* rational beings, you also imagine yourself to be *subject* to these laws—after all, you are a rational being yourself. The question you are to ask concerning any specific maxim is whether you can will to act on the maxim and simultaneously will that it be a law for all, including yourself (GMS 4:435–436). If

[2] I do not summarize the full argument here. For example, I do not touch on the analogy between the Categorical Imperative and a state constitution (cf. Kleingeld, 2018a, 166–168), since the comments do not call that part of my argument into question.

you cannot, because this would involve a contradiction, then acting on the maxim is impermissible.[3]

The Formula of Autonomy's procedure for assessing the moral status of maxims parallels Kant's criterion for the justice of state laws as explicated in the *Feyerabend Lectures on Natural Law*. In these lectures, held during the months in 1784 when Kant was writing the *Groundwork*, he argues that the laws of a state are just if they could stem from the agreement of the people as a whole. He explains that this requires that laws be genuinely universal, allowing no exceptions in the service of private interests. For example, laws including exceptions for the ruler's favorites fail to meet this test (V-NR/Feyerabend 27:1382).

The ruler need not ask whether the people *actually* endorse a particular law or *would* endorse it when given the opportunity (V-NR/Feyerabend 27: 1382). Indeed, Kant reportedly told his students that 'it is not necessary for [the ruler] to judge whether the people *would* make such a law …, but whether it *could have* made such a law' (V-NR/Feyerabend 27:1382, emphasis added). The mere *possibility* of universal agreement suffices for a law to count as just: 'One must examine whether the law *could have* arisen from the agreement of all: if so, then the law is right' (V-NR/Feyerabend 27:1382, emphasis added).

In legislating accordingly, the highest legislator acts as the 'representative' of the people; the people are sovereign in the idea (V-NR/Feyerabend 27:1382). But nowhere does Kant argue that the legislator must be elected by the people.[4] He instead seems to speak of an unelected sovereign ruler or 'autocrat' (V-NR/Feyerabend 27:1382) who is so enlightened as to examine whether the laws he considers enacting could arise from the agreement of his people.

In the *Metaphysics of Morals*, the Formula of Autonomy is missing; moreover, Kant hardly uses the term 'autonomy' anymore. I suggest (in Kleingeld, 2018b) that the explanation for the disappearance of the formula may be found in the fact that Kant's normative account of political legislation underwent significant changes between 1784 and the mid-1790s. In *Toward Perpetual Peace* (1795) and the *Metaphysics of Morals* (1797), Kant argues that just legislation requires not only that the united citizens *could* agree, but also that they *do* agree. Citizens are members of a state who are 'united for the purpose of legislation' (MS 6:314). The freedom of a citizen, Kant now argues, is the authority to obey no external laws other than those to which he '*has* given' consent (MS 6:314, emphasis added). He describes the ('active') citizen as a 'co-legislator' [*Mitgesetzgeber*] (MS 6:335; cf. 345). 'Active' citizens have the 'right to vote' and the right 'to cooperate for introducing certain laws', in addition to other political rights (MS 6:315). 'Passive' citizens lack these rights but are subject to and protected[5] by the law. Indeed, Kant describes the active citizens themselves as voting for laws ('the positive laws, for which they vote', MS

[3] For further discussion of the simultaneity condition and the nature of the relevant contradiction, see Kleingeld, (2023) in this issue.

[4] There is nothing in the notion of 'representation' as such that requires that representatives be elected. For example, for centuries men represented women (in court, contracts, and so on) without having been elected to do so.

[5] This term sounds like a euphemism, given that the law (as conceived by Kant) denies them crucial civil rights. What is meant here is that they are not outlaws.

6:315). The model here is clearly no longer that of the unelected autocrat who gives the people laws they *could give* themselves. Instead, Kant presents the active citizens themselves as *co-legislating* 'by means of' their elected representatives (more on which below).

Whereas Kant changed the criterion for *just political laws*, he continued to articulate the criterion for *morally right maxims* in terms of *possible* universal legislation, that is, as the requirement that one's maxim *be able* to hold simultaneously as universal law (MS 6:225). As a result, in the *Metaphysics of Morals* the criterion for the justice of political laws no longer provides a structural parallel by which to explicate the criterion for the moral rightness of maxims. In other words, the two criteria no longer have a parallel structure such that the first can serve as the basis for an analogy by which to present the second (on Kant's notion of analogy, see Kleingeld, 2018a).

Kant's terminology in *the Metaphysics of Morals* now suggests a somewhat different political analogy for presenting the moral criterion, one that draws a parallel with the moment when citizens evaluate whether a legislative proposal *qualifies* for legislation. Before they enact a law, they need to determine whether their legislative proposal satisfies the formal prerequisites for legislation. At *this* point citizens should ask whether the proposed law *could* be adopted by all citizens together, that is, whether it has the required characteristic of genuine universality. Accordingly, Kant now formulates the moral requirement in terms of a maxim's *qualifying* as a general law (MS 6:225, 389, 393, 451). Because enlightened autocratic rulers should also ask this question before legislating, the terminology of 'qualifying' for universal legislation is found in Kant's earlier work as well (*qualificiren*, KpV 5:27; see also *tauglich, schicklich*, and cognates in GMS 4:438, 441, 444; KpV 5:27–28, 36, 74). Thus, although Kant drops one of the political analogies with which he presents the moral criterion (dropping the Formula of Autonomy), he does not change the substance of the moral criterion.[6]

3 Baiasu and Newhouse's Continuity Defense

In their comments, Baiasu and Newhouse object that even if Kant's political theory underwent the changes I describe, these changes do not make the original analogy inapt. From the 1780s through the 1790s, they argue, Kant held that laws should be given by the people's *representatives*, who are to give laws that the people as a whole *could* adopt and as such can *regard* as *their own*. As Baiasu puts it, *qua* representation relation, 'the relation of the ethical agent to the highest legislator is not different

[6] In this section I restate the argument of my earlier papers. I would here like to add a comment on a phrase I have since noticed in the *Metaphysics of Morals* and which might be read as having an affinity with the Formula of Autonomy. To find out whether your maxim *qualifies* as universal law, Kant remarks, you should 'think' of yourself as giving universal law (MS 6:225). Kant does not disambiguate whether you should think of yourself as a co-legislator or as an autocrat, however; in light of his theory in the Doctrine of Right, it could well be the former. Moreover, the formula Kant offers here is 'act in accordance with a maxim that *can simultaneously hold* as a universal law!', not the Formula of Autonomy, which has disappeared.

from her relation to the elected representative' (Baiasu 2023). After all, both involve legislation by representatives.

Newhouse adds that whether the people's representatives acquired their legislative authority 'by election, heredity, or conquest' is irrelevant since this is 'temporally upstream' from the representation relation (Newhouse 2023). If it does not matter how representatives came to occupy their role, then the introduction of *elected* representatives in Kant's later theory cannot explain the disappearance of the Formula of Autonomy in the 1790s.

Newhouse further argues that the Formula of Autonomy could also be read as demanding that the moral agent take up the perspective of the *general will* rather than that of an enlightened autocrat. In that case, the formula was never tied to the perspective of the enlightened autocrat to begin with. This would further strengthen the case for the formula's 'enduring validity' (Newhouse 2023).

In sum, in their thoughtful comments, Baiasu and Newhouse put pressure on my explanation of the disappearance of the Formula of Autonomy by arguing that even if Kant did abandon the formula, 'he did not need to' (Baiasu 2023).[7]

4 Three Different Representation Relations

I would like to address Baiasu and Newhouse's main line of objection by arguing that the representation relation in the *Feyerabend* lectures is different from that in the *Metaphysics of Morals*. Certainly, Kant describes both the unelected enlightened autocrat and the elected delegates in parliament as 'representing' the people. But the *representation relation* changes radically, or so I argue.

There are many different models of political representation (see Dovi, 2018) and, accordingly, different ways in which representatives can be involved in giving law. For the purposes of the present discussion, I will distinguish three types of representatives and their respective roles in just legislation, before determining which one(s) Kant endorses. For the distinction between the 'trustee' and 'delegate' models of representation, I draw on Hanna Pitkin's influential discussion (Pitkin, 1967, esp. 112–143); for the normative principles, I draw on Kant. The three types are as follows:

(1) the *unelected enlightened autocrat*: the enlightened autocrat who conquers or inherits the throne and is the representative of the people, whether self-appointed or by law. This autocrat decides which laws to enact, choosing only from the set of laws that the people as a whole *could* adopt (hence 'enlightened' autocrat). The people do not themselves decide which laws are enacted.

(2) the *elected trustee*: the elected representative to whom citizens outsource the business of legislating, usually for a specified period (until the next elections). Trustees legislate independently as they see fit, but they are bound to give only laws that

[7] In a recent article, Stefano Bacin argues for the stronger thesis that even though 'no passage in the *Metaphysics of Morals* alludes to the Principle of Autonomy', Kant 'still advocated' the principle (Bacin, 2022, 92–93). The textual evidence he uses to support his claim, however, is from the Vigilantius lectures, held in 1793-94, that is, well *before* the *Metaphysics of Morals* (Bacin, 2022, 93–95). Therefore, this evidence does not show that Kant still advocated the principle in the *Metaphysics of Morals*.

the citizens *could* adopt. As far as the legislative process goes, the trustees decide *for* the citizens; the citizens do not themselves decide which laws will be enacted.[8]

(3) the *elected delegate*: the elected representative who acts as the instrument of citizens in the process of legislation. The citizens decide which laws are enacted, selecting only from among the proposals that qualify for genuinely universal legislation; they legislate 'by means of' their representatives by giving them specific instructions as to how to vote on legislation.

On the first model, the subjects neither elect their representative nor decide which laws to enact; the autocrat decides. On the second, the citizens elect their representatives; the latter decide which laws to enact. On the third model, the citizens elect their representatives and decide which laws to enact; the representatives act as their instruments.

The practical difference between the second and third models can be illustrated by the system of political representation practiced in the Dutch Republic until 1795. Delegates went to the Estates General with strict instructions; they had little or no discretion. If the terms of proposed agreements changed during deliberations, then the delegates had to go back to their province and obtain new instructions before they could vote.[9] These delegates must carry out the will of their constituents. But the delegate model need not involve such case-by-case instruction. When delegates are elected on the basis of concrete party programs, for example, they need not consult their constituents repeatedly, since they carry out the will of their voters when making laws in accordance with their party program. Either way, the situation of such delegates is very different from that of elected trustees, who have the authority to decide independently and make laws for the people as they see fit, and who may have been elected on the basis of their personal qualifications rather than a political platform.

Of course, these three models are idealized positions on a spectrum, and there are many intermediate possibilities. Elected representatives can be granted more or less discretionary authority. For example, elected representatives in a political party system may be granted discretionary authority within parameters set by a party program that provides general directions rather than specific plans. Such a system combines elements of the second and third models.[10]

In the *Feyerabend* lectures, Kant defends the first model. In the *Metaphysics of Morals*, he seems to defend the third. He does not describe the elected representatives in terms of the trustee model, that is, as legislating independently 'for' or 'on behalf of' the citizens, and as giving laws the people *could give* themselves. As indicated above, he rather characterizes the ('active') citizens as *giving* laws themselves. He describes them as co-legislators, as voting for positive laws, and as giving (or having

[8] Arthur Ripstein seems to read Kant as advocating the trustee model of representation, on the basis of textual evidence from before 1795. For discussion, see Kleingeld (Forthcoming).

[9] In the Dutch system the delegates were representing provinces, not individual citizens, but this makes no difference for the purposes of explaining the *structure* of the representation relation.

[10] Further, as Newhouse rightly observes, in the *Feyerabend* lectures Kant observes that monarchs can gain their title either by birth, by election, or by a combination of the two (V-NR/Feyerabend 27:1388, Newhouse 2023). A monarch elected by the people would fit the trustee model of representation. Kant here proceeds to mention two *disadvantages* of electing monarchs and does not advocate the practice (V-NR/Feyerabend 27: 1388–1389).

given or having been able to give) consent to legislation.[11] He writes that the functions of the republic, which include legislation, should take place 'by all citizens united, by means of (*vermittelst*) their delegates' (MS 6:341). These and other formulations mentioned in Sect. 1 do not suggest that he conceives of citizens as *outsourcing* legislation to their representatives (as on the trustee model). They rather suggest that he conceives of them as *legislating* themselves, *by means of* their delegates (as on the delegate model).

My reply to Baiasu and Newhouse, therefore, is that the *relation* between the representative and those they represent changes fundamentally. According to the *Feyerabend* lectures, the enlightened unelected autocrat is to decide independently which laws to give to the people as a whole (choosing from among the laws that satisfy the formal requirements); this is the model Kant uses in the Formula of Autonomy. According to the *Metaphysics of Morals*, by contrast, the citizens themselves are to decide jointly which laws to enact, acting through their elected delegates in parliament. This change is also reflected in the wording of the criterion for just legislation, which Kant initially describes in terms of laws the people *could* give themselves, and later in terms of laws the citizens *have* agreed to. This change of wording is hard to explain if we assume, with Baiasu and Newhouse, that the representation relation in the *Metaphysics of Morals* is no different from that in the *Feyerabend* lectures.

As for Kant's Doctrine of Right in the *Metaphysics of Morals*, I have focused on the relation between the active citizens and their representatives, since this is the relation that is central to Kant's discussion and hence also to the comments by Baiasu and Newhouse. Kant devotes significantly less philosophical attention to the relation between active and passive citizens. The latter lack the right to vote and to take active part in the affairs of the state. Kant includes in this category all women, all children, and all men whom he saw as lacking 'civil independence', such as servants, day laborers, and the unemployed (MS 6:314–315). In other words, passive citizens make up the vast majority of the population. Importantly, the relation between active and passive citizens parallels the relation between the enlightened autocrat and their subjects: the active citizens give laws *to* the passive citizens without the latter's having elected them to do so, and they should give laws that *could* be adopted by the people as a whole. Kant fails to address the resulting tensions in his account of political freedom and citizenship, although he does acknowledge that the very notion of passive citizenship raises questions (MS 6:314).

5 'Autonomy' in the Doctrine of Virtue, and Why Kant Abandoned the Idea of Autonomy of the Will

In Kleingeld (2018b) I observed that the *Formula* of Autonomy disappears in the 1790s and that the very *terminology* of 'autonomy' recedes into the background as well. Despite its centrality in the *Groundwork*, the term does not occur in the theo-

[11] This need not mean that each individual citizen must consent to each law. In the *Metaphysics of Morals* Kant fails to make explicit, however, whether he sees unanimity as required or whether a majority decision counts as a decision of the united citizens.

retical set-up that Kant provides in the Introduction to the *Metaphysics of Morals*. I also pointed out, however, that Kant still uses 'autonomy' twice in the Doctrine of Virtue, and I added a brief discussion of the passages at issue (Kleingeld, 2018b, 77–78; MS 6:383, 480). Baiasu wonders whether the fact that Kant continues to use the term—only sporadically, but still—poses a problem for my reading (Baiasu 2023, n.9). I would therefore like to discuss the relevant passages in more detail and offer a possible explanation of why Kant abandoned the notion of the autonomy of the will.

In the *Groundwork*, Kant calls autonomy a property of the *will*. He argues that the moral laws to which the will is *subject* (*unterworfen*) must be viewed as *the will's own legislation* (GMS 4:432). Autonomy, he writes, is 'the property of the will by which it is a law to itself' (GMS 4:440). The will (*Wille*) here appears in the role of both legislator and subject. Throughout this discussion, Kant uses the word '*Wille*' (dozens of times), using *Willkür* only twice and without a clear difference in meaning (GMS 4:428, 451).

In the Doctrine of Virtue of the *Metaphysics of Morals*, by contrast, Kant no longer ascribes autonomy to *Wille*. Here, he explicitly distinguishes two volitional capacities, *Wille* and *Willkür*. He now identifies *Wille* with practical reason and the faculty of moral legislation, and he identifies *Willkür* with the source of maxims and the faculty of deciding whether and how to act (MS 6:213, 226). He writes that pure practical reason (*Wille*) gives laws '*for*' *Willkür* (MS 6:214). Thus, the legislator and the subject are no longer identical; *Wille* can no longer be said to be a law *to itself*.

The fact that Kant no longer considers *Wille* to be 'a law *to itself*' arguably explains why he abandons his earlier conception of autonomy of the will. After all, in the *Groundwork* he had defined autonomy precisely as the will's property of being 'a law to itself'. Relatedly, when in the Doctrine of Virtue Kant again uses the word 'autonomy', it must either mean something different than it did in the *Groundwork* or be predicated of another entity. I therefore now turn to the two passages in which the word occurs to determine whether Kant uses 'autonomy' in a different sense or predicates it of an entity other than *Wille*.

The first thing to note is that in both cases Kant speaks of the 'autonomy *of practical reason*'. The meaning of this phrase is not immediately clear. Since reason as such is not *subject* to moral laws, its autonomy cannot consist in its being subject to its own laws. One might perhaps assume that Kant is here using 'autonomy' in a different sense, as referring merely to the legislative power of reason. In support of this assumption, one could point to a contemporaneous book by Johann Christian Maier, who defined autonomy as 'the power to legislate on one's own authority' (*die Eigenmacht, Gesetze zu geben*, Maier, 1782, 3).[12] Maier saw autonomy as including the power of legislating *for others*. In line with Maier's understanding of the term, the 'autonomy of practical reason' could then be taken to refer to the fact that practical reason, on its own authority, gives moral laws to human beings (or to their *Willkür*). However, as an early reviewer of the book pointed out, to the extent that Maier takes autonomy to mean the power to give laws to others, this definition is implausibly far removed from the common meaning of the term (Fk [Runde] 1782, 409).

[12] I thank Frederick Rauscher for drawing my attention to Maier's book.

It seems more plausible to assume that Kant, in the Doctrine of Virtue, predicates autonomy of an entity other than the will, namely the human moral subject as a whole, including both *Wille* (practical reason) and *Willkür*. Human beings are both *subject* to moral laws and subject to *their own legislation*, since moral legislation stems from practical reason, and practical reason is their own faculty. This assumption seems to fit the passage where Kant refers to the 'subjective autonomy of the practical reason of each human being' (MS 6: 480). Here, Kant emphasizes that moral laws, given by practical reason, ought to serve humans as incentives. The expression 'the autonomy of practical reason' would indicate not that *practical reason* is subject to its own laws but that the moral laws to which each *human being* is subject *stem from their own reason*. Similarly, the expression 'the writings *of the philosopher*' may (and usually does) refer to writings the philosopher has produced. That is, autonomy would be a property not of practical reason (*Wille*) but of human beings, and it would be the autonomy 'of reason' in the sense that their own reason is the source of the moral laws to which they are subject.

Second, this assumption may also fit the difficult passage where Kant describes a 'doctrine of morals' for finite holy beings as 'an autonomy of practical reason' (MS 6:383). Finite holy beings can regard moral laws as stemming from their own reason. As for humans, who are not holy beings, Kant goes on to say – perhaps a bit provocatively – that a 'doctrine of virtue' for humans is not just 'autonomy' but also, at the same time, 'autocracy' of reason: in humans, reason should have not only legislative but also executive authority (MS 6:383).[13] Perhaps, then, the fact that the relevant 'self' is broadened from the will to the human moral subject as a whole may again explain why reason's legislation can be described as 'autonomy': moral laws are a human being's *own* laws, in the sense that they stem from one's own reason.[14]

If this is right, it fits with the above explanation for why the term 'autonomy' lost its prominent status. In both passages, the self that is subject to the moral laws (viz., the human being, by virtue of its faculty of choice) is no longer identical to the self that legislates (viz., practical reason). As a result, the term 'autonomy' is potentially misleading. It could mistakenly be understood as implying that human beings as such give themselves moral laws, and then Kant could be misread as being a moral

[13] With reference to this passage, Bacin argues that 'autonomy is there considered not primarily as reason's self-legislation' but is 'used there precisely to denote autocracy', that is, as reason's self-government in the maxims of rational agents (Bacin, 2022, 98). I do not see Kant as equating autonomy and autocracy in this passage, however. Kant mentions that a doctrine of morals (*Sittenlehre*) 'is an autonomy of reason', and here clearly reason's *legislative* authority is meant as a source of moral laws. A 'doctrine of virtue', Kant adds, is 'at the same time an autocracy' (MS 6:383). 'At the same time an autocracy' here suggests that in a doctrine of virtue reason has supreme legislative *and* executive authority: reason is not just the source of moral laws but *also* the power to master one's rebellious inclinations (MS 6:383). Thus, contrary to Bacin's assertion, 'autonomy' does not seem to be used to denote autocracy; autonomy and autocracy are distinct aspects.

[14] Alternatively, the broadened 'self' could be taken to be the higher faculty of desire, which includes both *Wille* and *Willkür*. If we take the relevant self to be the higher faculty of desire, the legislator and subject would again not be identical, since *Wille* would legislate for *Willkür*. I take the broadened 'self' to be the human being as a whole rather than the higher faculty of desire since in the two passages under consideration Kant mentions the human being, not the faculty of desire. But it is clear that autonomy can be ascribed to the human being only by virtue of its higher faculty of desire. I thank Micha Gläser for raising this issue.

voluntarist. The history of Kant interpretation confirms that this has indeed happened repeatedly (Wood 2008: 106–11). In the initial account of the *Groundwork*, the non-identity of the legislator and the subject of moral laws had not become clear yet, as Kant used the term *Wille* to refer to both. In the *Metaphysics of Morals* Kant thematizes it explicitly by introducing the distinction between *Wille* and *Willkür* and asserting that the former gives laws for the latter.

Although the conception of autonomy that is operative in these two passages certainly merits further discussion, I believe it is clear enough that it is *discontinuous* with his *Groundwork* theory of autonomy as the will's property of being a law to itself. Yet the relevant passages also make clear that Kant has not given up the *core idea* he expressed in that earlier theory, namely that the origin of morality must lie in pure practical reason.

6 Two Further Objections

Baiasu and Newhouse raise two further issues, both in support of the presumption of continuity in Kant's position. These concern his possible reference to the 'general will' in the *Groundwork* and the role he attributes to God as an ethical legislator.

6.1 The Ruler or the United General Will?

Newhouse argues that the perspective the moral agent should adopt according to the *Groundwork*'s Formula of Autonomy is perhaps not best understood as that of a sovereign legislator giving laws to all. Rather, she suggests, 'Kant may well have intended to compare a human will to the united general will of the people itself' (Newhouse 2023) rather than the will of an enlightened autocratic legislator. If Kant's phrase '[*eines*] *allgemein gesetzgebenden Willens*' (quoted at the beginning of Sect. 1 above) is translated as '[of a] universally-legislative will', she argues, this could perhaps—Newhouse calls this an 'intriguing suggestion'—refer to the 'united general will' (Newhouse 2023). This possibility is obscured by translations that turn the phrase into an adverbial clause ('[of a] will giving universal law').

The significance of this issue, I take it, is as follows. If the Formula of Autonomy in the *Groundwork* is read as enjoining the agent to adopt the perspective of the *general will*, then the fact that Kant had second thoughts on the appropriateness of enlightened *autocracy* is simply irrelevant to the Formula of Autonomy and cannot explain its disappearance. This would make room for the possibility that Kant remained (albeit tacitly) committed to the formula.

I agree with Newhouse that the phrase itself could be translated in the way she suggests. The reason I turned it into an adverbial clause, as Gregor also did, is that the original German word order yields an ungrammatical result in English. Kant would be speaking of 'the idea of the will as a through all of its maxims universally-legislative will'—so some change seemed needed. Newhouse prefers adding parentheses around 'through all of its maxims'.

Even if we were to adopt Newhouse's translation, however, I doubt that the phrase 'universally-legislative will' refers to the general will. To my knowledge, there is no

passage in the *Groundwork* where Kant connects *allgemein* and *Wille* so as to refer to the 'general will'. In both passages quoted by Newhouse, as well as in the subsequent sentence, *allgemein* ('universally' or 'generally') belongs to *gesetzgebenden*, not to '*Willens*'. The expression 'universally legislating' and cognate expressions ('universal law', etc.), without reference to 'will', are found roughly forty times in the relevant pages alone (GMS 4:421–424; 4:431–435). Furthermore, several versions of the formula refer to the moral subject's own will rather than a universally-legislative [general] will. For example, the statement that each rational being 'must regard *himself* as legislating universally through all the maxims of *his will*' (GMS 4:433, emphasis added) cannot be read as referring to the general will. In sum, I believe the wider context suggests that the 'will' mentioned in the Formula of Autonomy does not refer to the general will.

6.2 God's Ethical Legislation

In 2018b I added, as a subsidiary comment, a reference to the passage in the *Religion* in which Kant argues that in a 'juridical commonwealth' the people as a whole 'is itself the legislator', whereas in an 'ethical commonwealth' 'the people as such cannot itself be regarded as legislating' (RGV 6:98). In the ethical commonwealth, only God can be represented as legislating. God is conceived as a moral ruler of the world who legislates only 'genuine duties', that is, duties grounded in pure reason (RGV 6:99). Baiasu argues that this passage does not help my case. The reason why Kant introduces God here is that ethical legislation is a matter of inner motivation, Baiasu argues, and only God knows the inner disposition of agents. This is a point that Kant already made in the *Groundwork*, as he (rightly) points out, so in his view it cannot explain the disappearance of the Formula of Autonomy (Baiasu 2023).

My reason for drawing attention to this passage, however, was not to suggest that Kant changed his view on ethical legislation—quite the opposite. My point was rather that he changed his view on *political* legislation and that, as a result, he could no longer consider the ethical commonwealth to be analogous to a juridical commonwealth with regard to legislation. The passage in the *Religion* confirms this point by presenting the realms of political and moral legislation as structurally *disanalogous*, whereas in the *Groundwork* he had still presented the moral realm and the political realm as analogous.

In the *Groundwork*, Kant presented three relations as having a parallel structure: the enlightened autocrat giving just laws to his people; God (represented as) giving moral laws to the realm of ends; and you (acting as if you are) giving laws to all rational beings, as per the Formula of Autonomy. In the 1790s, this changes, as is illustrated by the passage from the *Religion*. Here, Kant still writes that God is represented as giving moral laws to the ethical commonwealth, but he now also asserts that in the juridical commonwealth the people itself legislates.

7 Caution: Merging Terms—from *Sittlich, Moralisch*, and *Ethisch* to *Moral* and *Ethical*

In this final section, I would like to address a very different point raised by Baiasu. It is not directly related to the issue of the disappearance of the Formula of Autonomy, but it certainly merits discussion. He helpfully explains his choice of terminology for describing the distinction between external and internal lawgiving. Baiasu calls the first 'political' and, taking his cue from the *Metaphysics of Morals*, he calls the second 'ethical' legislation (Baiasu 2023). In the two essays to which Baiasu and Newhouse respond, I use 'legal and political' (or 'political') and 'moral' instead, and I would like to take this opportunity to explain why I use 'moral' rather than 'ethical'. This is not meant as a criticism of Baiasu's usage; there are different ways to deal with what is in effect an impossible situation.

What makes the situation impossible is, first of all, the fact that there are *three* German terms, namely, *sittlich*, *moralisch*, and *ethisch* (with Germanic, Latin, and Greek roots, respectively), but *only two* available English translations, namely 'moral' and 'ethical'. *Ethisch* is simply translated as 'ethical', but both *sittlich* and *moralisch* are usually translated as 'moral',[15] even though, for Kant, they are not synonyms.

To make things even more complicated, Kant's use of the three terms and their cognates (e.g., *Sitten*, *Moral*, *Ethik*) varies. For example, in the Mrongovius lectures on moral philosophy, Kant explains that *Ethik* can be used in a broad and narrow sense. In the broad sense, it refers to practical philosophy as a whole. In the narrow or 'proper' sense, it is synonymous with 'doctrine of virtue' (V-Mo/Mron 29:630). In the Preface to the *Groundwork*, Kant uses *Ethik* in the broad sense, indicating that it is synonymous with *Sittenlehre* and includes both an empirical and a rational part, the latter of which he calls *Moral* (GMS 4:387–388). In the *Metaphysics of Morals*, by contrast, he uses *Ethik* in the narrow sense, as referring to only to the Doctrine of Virtue.[16] As a result, if we use 'ethical' in the narrow sense of the *Metaphysics of Morals* to translate *sittlich* or *moralisch* in the *Groundwork*, we run the risk of unduly narrowing the meaning of these terms. Moreover, Kant does not use *ethisch* in the *Groundwork*, and 'ethical' is not an obvious choice for translating *sittlich* or *moralisch.*

It is beyond the scope of this essay to discuss these and similar difficulties in detail, but the upshot is that it is impossible to use the English words 'moral' and 'ethical' in a way that fully matches Kant's use of the three German terms. Or, put differently, there is no way to capture the full complexity of Kant's usage of the three German terms by imposing any strict rules on the use of the two available English translations. Current Anglophone authors will have to be creative, use terms in their current sense, and/or provide the necessary exegetical and semantic background information. I hope (and believe) that the expressions 'moral and political' and 'moral and legal',

[15] See, for example, the *Groundwork* translations by Mary Gregor and Allen Wood. Gregor makes this explicit concerning her translation of the related nouns: '*Moral* and *Sitten* are translated as "morals," *Moralität* and *Sittlichkeit* as "morality"' (note c to GMS 4:389). Allen Wood makes this explicit in his glossary.

[16] Accordingly, Kant distinguishes between '*Moral als Ethik*' and '*Moral als Rechtslehre* [Doctrine of Right]' (ZeF 8: 383–384). Here, *Moral* refers to the philosophical doctrine.

when read in context, sufficiently indicate to readers which distinctions are meant. The use of these distinctions seems justifiable in terms of Kant's continued use of the distinction between 'morality' and 'legality' in the *Groundwork* and his distinction between 'moral' and 'political'.[17] Given the complexity of the situation, however, there is ample room for others to make different choices.

8 Conclusion

I am deeply grateful to Baiasu and Newhouse for their thoughtful comments, which required me to clarify and further develop my argument. I have offered additional support for my original contention that the analogy underlying the *Groundwork*'s Formula of Autonomy became inapt when Kant changed his views on political representation and on the criterion for just legislation. I have also suggested a possible explanation of the disappearance of autonomy as a property of the will. While Kant's position in the *Metaphysics of Morals* is discontinuous with the account in the *Groundwork* in important respects, I have also emphasized that Kant did not change his view on the *substance* of the moral criterion or on its assumed *origin* in pure practical reason.

Acknowledgements I thank Sorin Baiasu for organizing the 2018 KOSAK conference at the University of Keele, where he and Marie Newhouse first presented their comments, and I thank him and Micha Gläser for editing this special issue. I also thank them, as well as Suus Jacobi, Marie Newhouse, Vinicius Pinto de Carvalho, Janis Schaab, Fiorella Tomassini, and Marijana Vujošević for helpful comments and discussion. I thank the Dutch Research Council (NWO) for financial support.

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

References

Baiasu, S. (2023). How Far Do We Self-Legislate? *Philosophia.* https://doi.org/10.1007/s11406-023-00613-4.

Bacin, S. (2022). The Fate of Autonomy in Kant's *Metaphysics of Morals. British Journal for the History of Philosophy*, *30*, 90–108.

Dovi, S. (2018). Political Representation. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/fall2018/entries/political-representation/.

[17] I do not mean to imply that I have always been consistent in my use of these terms.

Fk [Justus Friedrich Runde]. (1782). Review of Johann Christian Maier, *Autonomie, vornemlich des Fürsten- und übrigen unmittelbaren Adelstandes im R[ömischen] deutschen Reiche*. *Allgemeine Deutsche Bibliothek*, *50*, 408–410.

Kant, I. (1996). *Practical Philosophy*. Translated and edited by M. J. Gregor. Cambridge University Press. https://doi.org/10.1017/CBO9780511813306.

Kleingeld, P. (2018a). Moral Autonomy as Political Analogy: Self-legislation in Kant's *Groundwork* and the *Feyerabend Lectures on Natural Law.* In S. Bacin and O. Sensen (eds.), *The Emergence of Autonomy*. Cambridge: Cambridge University Press, 158–175.

Kleingeld, P. (2018b). The Principle of Autonomy in Kant's Moral Theory: Its Rise and Fall. In E. Watkins (ed.), *Kant, Persons, and Agency* (pp. 61–79). Cambridge: Cambridge University Press.

Kleingeld, P. (2023). A Defense and Development of the Volitional Self-Contradiction Interpretation. *Philosophia.* https://doi.org/10.1007/s11406-023-00641-0.

Kleingeld, P. (Forthcoming). Independence and Kant's Positive Conception of Freedom. In M. Timmons and S. Baiasu (eds.), *Kantian Citizenship: Grounds, Standards and Global Implications*. London: Routledge.

Maier, J. C. (1782). *Autonomie, vornemlich des Fürsten- und übrigen unmittelbaren Adelstandes im R[ömischen] deutschen Reiche*. Tübingen: Heerbrandt.

Newhouse, M. (2023). The Principle of Autonomy's Enduring Validity. *Philosophia.* https://doi.org/10.1007/s11406-023-00638-9.

Pitkin, H. F. (1967). *The Concept of Representation*. Berkeley: University of California Press.