CHAPTER 4

The Principle of Autonomy in Kant's Moral Theory: Its Rise and Fall

Pauline Kleingeld

1 INTRODUCTION

The notion of autonomy is absolutely central to Kant's moral theory in the *Groundwork* and the *Critique of Practical Reason*. Kant considers autonomy of the will to be the "supreme principle of morality" (G 4:440) and the "sole principle of all moral laws and the duties corresponding to them" (KpV 5:33). Kant presents the "Principle of Autonomy" (known in the literature as the "Formula of Autonomy") as an especially apt version of the Categorical Imperative (G 4:431–2), describing it as the "principle of each human will as a will that is legislating universally through all of its maxims" (G 4:432). Furthermore, he calls autonomy a "property" of the will (G 4:440), namely the property of its being the source of the laws to which it is subject, independently of inclination or any other authority outside the will itself. Last but not least, he equates autonomy with freedom of the will (G 4:447). Clearly, the idea of autonomy is of crucial importance.

By the time we get to the *Metaphysics of Morals*, however, 'autonomy' has virtually disappeared. In the Introduction, Kant does not mention it, let alone highlight it as the supreme principle of morality, even though this is where he lays out the basic concepts and presuppositions of the book. He reintroduces the Categorical Imperative, the notions of moral laws and duties, the idea of freedom, and many other core elements of his moral theory – but not autonomy. Indeed 'autonomy' occurs only twice in a moral context, each time without special emphasis and without reference to particular moral principles or freedom of the will (MS 6:383, 6:481). The Principle of Autonomy is not mentioned at all.

What happened? In the literature, there is no debate about the virtual disappearance of the notion of autonomy from *The Metaphysics of Morals*, and Kant himself does not comment on the issue. Most authors seem to assume that he was still committed to his previous views but simply failed to mention it. Given the book's stated aims, however, his failure to mention

this anywhere would be strange. Moreover, as I show below, autonomy recedes into the background well before the *Metaphysics of Morals*, and the Principle of Autonomy completely disappears.

If we wish to gain a better understanding of the curious fate of autonomy in Kant's ethics,[1] a good place to start is in the *Groundwork* where he first presents the notion of autonomy. This is his discussion of the third formulation of the Categorical Imperative, the formulation that he terms the "Principle of Autonomy." If it is possible to explain the rise and fall of this third formula, this may provide us with the key to explaining the virtual disappearance of 'autonomy' during the 1790s.

The thesis I shall be defending is that the Principle of Autonomy in the *Groundwork* includes a legislation analogy, and that this analogy was apt given the political theory Kant defended at the time, but that it became obsolete in the 1790s, when he made important changes to his political theory. Around the time of the *Groundwork*, Kant held that for state laws to be fully just, it is sufficient that they be genuinely universal; it is not necessary that the citizens actually consent to the laws. This made it possible for Kant to regard the criterion governing the moral permissibility of maxims as being fully analogous to the criterion governing the justice of political legislation. Thus, he could express the principle of morality as the requirement to act only on maxims that one can simultaneously regard oneself as *legislating universally*. During the early 1790s, however, Kant dropped the view that genuine universality suffices for the justice of state laws and added a second requirement, namely that the laws be given by the citizenry. He did not make a parallel change to his moral theory. As a result, I argue, the idea of "*legislating* through one's maxims" was no longer suitable as an analogy with which to express the moral principle. This explains why Kant no longer mentions the Principle of Autonomy.

In what follows, I first discuss the legislation analogy as Kant presents it in the *Groundwork* and the *Critique of Practical Reason*. I examine Kant's exposition of the idea of autonomy, emphasizing, among other things, that it involves legislation to all rational beings, including oneself – not legislation primarily "to oneself" (Section 2). I then explain the legislation analogy in further detail with reference to Kant's 1784 political theory (Section 3). I next explain, in light of the changes Kant makes to his political theory

[1] Karl Ameriks wrote a masterful book outlining the "fate of autonomy" after Kant in the works of Reinhold, Fichte, Hegel, and others. See Karl Ameriks, *Kant and the Fate of Autonomy: Problems in the Appropriation of the Critical Philosophy* (Cambridge: Cambridge University Press, 2000). I would like to suggest that there is also an interesting story to be told about the fate of autonomy in Kant's own work – in particular, about its rise in the *Groundwork* and its virtual disappearance from his moral theory in the 1790s.

in the 1790s, why the idea of autonomy no longer provides a fitting analogy with which to express the principle of morality. I also consider what takes its place, namely the idea of a maxim's *qualifying* as a universal law (Section 4). In the final section, I discuss the two passages in the Doctrine of Virtue in which Kant speaks of autonomy as a property of practical reason, in order to show that they fit with the account provided in this chapter. I argue that neither passage refers to the Principle of Autonomy as a formulation of the Categorical Imperative (Section 5).[2]

2 MORAL AUTONOMY AS A POLITICAL ANALOGY

The notion of autonomy of the will, which Kant first introduced in the *Groundwork*, includes the following two core elements: the idea that the will is subject to moral laws, and the idea that the obligatory force of these laws originates in the will itself. If the will's highest governing principles depended on an authority outside the will, this would be *heteronomy* – that is, the property of being subject to the legislation of another. Kant argues, however, that the will has the property of autonomy; it is subject to *its own* legislation, legislation that is independent of inclinations (e.g., G 4:432–3, 440).[3]

Kant not only describes autonomy as a *property* of the will but also formulates the so-called "*Principle* of Autonomy," which he presents as one of the main formulas of the Categorical Imperative.[4] Here, the idea of autonomy serves to formulate a procedure for examining whether one's maxims are morally permissible. Kant describes the Principle of Autonomy as the principle of regarding one's will as "a will that is universally legislating through all of its maxims" (G 4:431–2). He describes it in more detail as the command

> to do no action on any other maxim than so that it could also coexist with it [viz., with the maxim] that it be a universal law, and hence to act only so that the will could regard itself as simultaneously giving universal law through its maxim. (G 4:434, cf. 440)

[2] Another important issue in connection with the topic of this essay is the relation between the changes described here and the changes in Kant's theory of freedom of the will. In the *Groundwork*, Kant maintains that freedom of the will consists in autonomy, but he abandons this view – or at least changes it significantly – with the introduction of the distinction between *Wille* and *Willkür* in the *Metaphysics of Morals* (MS 6:213–14, 226). I defer discussion of this issue to another occasion, however, since it would require a chapter of its own.

[3] Paul Guyer articulates a somewhat different reading of autonomy, according to which it consists in the agent's capacity to control his inclinations and act from duty. See Paul Guyer, "Kant on the Theory and Practice of Autonomy," *Social Philosophy and Policy* 20 (2003): 70–98.

[4] On the distinction between autonomy as a property and as a principle, see also Henry E. Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), pp. 94–106.

Thus, in evaluating the moral permissibility of one's maxims, one is to imagine one's will as *legislating*. One ought to conceive of one's will as analogous to a political legislator who enacts state laws. This connection with political legislation is somewhat obscured by the fact that it is customary to speak of 'universal law' in the context of Kant's moral theory and of 'general law' in the context of his political theory, but the underlying German term is the same.[5]

Kant presents the Principle of Autonomy as involving the use of an *analogy*.[6] He does not argue that we *actually* give universal laws through our maxims. Rather, he states explicitly that several versions of the Categorical Imperative, of which the Principle of Autonomy is one, each involve the use of "a certain analogy" (G 4:436, cf. 437) or a different "way of representing" the Categorical Imperative (G 4:431, 436). The non-literal character of the Principle of Autonomy is manifest in Kant's repeated assertions that we are to "consider" or "regard" the will as simultaneously legislating through its maxims (see the passage quoted above, and G 4:431–4, 438) or that we are to "act as if" our maxims were to serve not merely as our individual action principles but simultaneously as laws for all members of the moral community, or "realm of ends":[7]

> [E]very rational being must act as if he were through his maxims at all times a lawgiving member of the universal realm of ends. The formal principle of these maxims is: Act as if your maxim were to serve at the same time as a universal law (of all rational beings). (G 4:438)

[5] In most cases, the German word '*allgemein*' is best translated as 'general,' but in English translations of Kant's moral theory it is usually rendered 'universal'. This is because Kant distinguishes between '*universelle*' and '*generelle*' rules (KpV 5:36, cf. also MS 6:216). The former rules hold always and necessarily, whereas the latter are merely correct on average and do not hold always and necessarily. '*Generell* ' is translated as 'general,' and '*allgemein*' as 'universal.' This is not a mistake, and in this chapter I shall do the same. It does obscure the connection, however, between 'general/universal laws' (*allgemeine Gesetze*) in morality and politics, since the laws of a state are usually called 'general' rather than 'universal.' Yet we are familiar with the use of 'universal' in some political contexts, for example in expressions such as 'universal suffrage' and 'universal healthcare.'

[6] This has also been noted in the Kant literature, for example in Andrews Reath, *Agency and Autonomy in Kant's Moral Theory: Selected Essays* (Oxford: Oxford University Press, 2006), Chapter 4, and Jens Timmermann, *Kant's Groundwork of the Metaphysics of Morals: A Commentary* (Cambridge: Cambridge University Press, 2007), pp. 110–11. Nevertheless, the fact that Kant speaks of an analogy is still often overlooked in discussions of Kant's notion of autonomy. I discuss Kant's conception of analogy in more detail in my essay "Moral Autonomy As Political Analogy: Self-legislation in Kant's *Groundwork* and the *Feyerabend Lectures on Natural Law*," in *The Emergence of Autonomy in Kant's Moral Theory*, ed. Stefano Bacin and Oliver Sensen (Cambridge: Cambridge University Press, 2018).

[7] For an overview of Kant's use of these locutions, see Allen W. Wood, *Kantian Ethics* (Cambridge: Cambridge University Press, 2008), p. 111. For a discussion of the importance of the simultaneity condition, see my essay "Contradiction and Kant's Formula of Universal Law," *Kant-Studien* 108 (1) (2017): 89–115.

Clearly, the Principle of Autonomy does not demand actual legislation in a literal sense but rather articulates a *counterfactual* procedure as a method for assessing the moral permissibility of maxims.

Two further clarifications are in order. First, the procedure Kant articulates does not require any actual or even imagined deliberation with others. He always presents the procedure as one that can be carried out by the individual agent entirely in thought, and at no point does he require that we imagine deliberation among the members of the "realm of ends."[8] The criterion for the moral permissibility of a maxim, articulated in the Principle of Autonomy, is whether this maxim can simultaneously be a universal law ("that it could also coexist with it that it be a universal law," G 4:434, quoted above). Although Kant mentions that we should take the perspective of all other rational beings into account (G 4:438), we satisfy this requirement simply by selecting maxims that can "serve at the same time as a universal law (of all rational beings)" (G 4:438). At no point in his discussion of the Principle of Autonomy does Kant claim that others' actual attitudes toward our maxims must be taken into consideration.

Second, Kant does not speak of giving laws "to oneself," not even analogically. He consistently speaks of giving *universal* law or of a law *of all rational beings*. The set of all rational beings includes oneself, of course, but we misrepresent the scope of the imagined legislation – or at least represent it misleadingly – if we describe the Principle of Autonomy as requiring that one act as if one were legislating only or primarily *to oneself*.

In current discussions, however, Kant is commonly read as saying just that: autonomy means that the will gives laws – or should "consider itself as" giving laws – *to itself*. On this reading, the scope of the law is broadened to other rational beings only mediately, via the idea that I give laws to myself *qua* rational being, and hence implicitly to all rational agents.

Importantly, however, Kant puts it precisely the other way around. The Principle of Autonomy requires that I conceive of myself, counterfactually, as giving universal law through my maxims – as legislating *to all* (including myself). The scope and addressee of the imagined legislation is the entire moral community: "all rational beings" (G 4:438) or "everyone who has

[8] In her article "Autonomy, Plurality and Public Reason," in *New Essays on the History of Autonomy: A Collection Honoring J. B. Schneewind*, ed. Natalie Brender and Larry Krasnoff (Cambridge: Cambridge University Press, 2004), pp. 181–94, Onora O'Neill rightly emphasizes that Kant presupposes a plurality of agents. He presents the idea of autonomy as the idea of the will "of *every rational being* as a will giving universal laws," and the Principle of Autonomy applies to a plurality of agents who ought to take the perspective of all others into account. Yet the moral imperative is addressed to each agent individually, and Kant assumes – as his *Groundwork* examples attest – that one can assess the moral permissibility of maxims entirely by oneself.

reason and will" (KpV 5:36). It is helpful to remember that Kant is using a political analogy here. It would be a misrepresentation to say that the role of a political legislator is to give laws *to himself*. Certainly, under the rule of law those who legislate are also subject to the laws they give, but legislators give laws *to all*, including themselves; they don't address the laws primarily *to themselves*.

There is of course a crucial element of reflexivity included in the idea of legislating to all, because universal laws are conceived as also applying to oneself. Kant emphasizes this by saying that the will is to be regarded as a legislating *member* of the moral community and, as such, subject to its own legislation (G 4:433). Yet one would misdescribe the will's imagined objective if one were to say that its aim is to give law to itself; its imagined objective, as Kant describes it, is to give universal laws – laws that apply to all rational beings.

In the whole of the *Groundwork*, there does not seem to be a single passage in which Kant writes explicitly and unequivocally that autonomy of the will consists in the will *giving law to itself* (except in translations, about which more below).[9] Rather, Kant speaks of considering the will as "giving *universal* law" and of this being the will's "*own* legislation" (e.g., G 4:431–2; KpV 5:33). Kant also speaks of the will's "*being* a law to itself" (G 4:440, 447). None of these expressions is synonymous with "*giving* law *to oneself*" in the sense of the *primarily self-addressing act*, on the part of the will, of *enacting* legislation.[10] Instead, Kant describes the imagined objective as giving laws to the entire imagined community (of which one is a member).

Why do these clarifications matter? If there is an undeniable element of reflexivity in the notion of autonomy, insofar as one considers the will itself to be subject to the laws that it gives to all, then why is it so important to specify precisely to whom the law is given? The reason is that it has implications for the role of consent. If the will is viewed as legislating *to itself*, its consent is implied. If, by contrast, the will is viewed as legislating *to all rational beings* – which is what Kant writes – then this raises

[9] The formulations that come closest are negative and concern heteronomy (G 4:444, KpV 5:33).

[10] There is one context in which Kant does explicitly use the reflexive expression in relation to laws, namely when he writes that reason gives laws to itself for thinking (*Gesetze* […] *die sie sich selbst gibt*, WDO 8:145). Indeed, if reason gives laws for thinking, it gives these laws to itself. But this does not in any way entail that the imagined universal legislation through our maxims, of which the Principle of Autonomy speaks, should also be conceived as legislation of the will "to itself," and Kant does not use the expression in this context.

questions concerning their (imagined) consent. In the next section, I discuss Kant's account of the role of consent in political legislation and show that this greatly illuminates his account of the Principle of Autonomy.

Before we move on to this topic, however, I must discuss one passage in more detail, as it is often thought to provide key evidence for the common interpretation. The impression that Kant claims that the will should be regarded as legislating *to itself* is due, it seems, to one particular word in one particular passage. This is the passage in which Kant writes – using a German equivalent of the Greek-derived word 'autonomy' – that the will must be viewed as "self-legislating" (*selbstgesetzgebend*).[11] In English translations, "*selbstgesetzgebend* " is often translated as "legislating to itself" (e.g., Allen Wood[12]) or as "giving the law to itself" (e.g., Henry Allison, Mary Gregor[13]). Since this is the only occurrence of this word in Kant's published work, determining its meaning requires a closer look at the text.

In the passage at issue, Kant articulates the Principle of Autonomy in terms of the will's *giving universal law*, its being *subject* to this universal legislation, and its being subject to this universal legislation *because* it is itself its "author":

> In accordance with this principle [viz., the Principle of Autonomy] all maxims are rejected that cannot coexist with the will's own universal legislation. The will is thus not merely subject to the law but subject in such a way that it must also be viewed as *self-legislating* [*selbstgesetzgebend*] and precisely for that reason subject to the law in the first place (of which it can regard itself as author [*Urheber*]). (G 4:431, emphasis in original)

As a matter of translation, 'self-legislating' seems the best choice because it preserves the ambiguity of the German original. But what does the word mean here?

There are at least four reasons for taking 'self-legislating' to indicate that the will is the *source* of legislation (that is, legislation *by* the will itself). First, instead of "*selbstgesetzgebend* " (one word), Kant also uses "*selbst gesetzgebend* " (two words), which simply means 'itself legislating.' For

[11] The Greek adjective αὐτόνομος, which derives from the words for 'self' and 'law', means 'independent' or 'living under one's own laws.'

[12] In his translation of the *Groundwork*, in Immanuel Kant, *Groundwork for the Metaphysics of Morals*, ed. and trans. Allen W. Wood (New Haven: Yale University Press, 2002).

[13] In Henry E. Allison, *Kant's Groundwork for the Metaphysics of Morals: A Commentary* (Oxford: Oxford University Press, 2011), p. 240; Mary J. Gregor in the *Practical Philosophy* volume of the *Cambridge Edition of the Works of Immanuel Kant.* In a footnote, however, Gregor mentions "itself lawgiving" as an alternative translation.

example, in the same context as the quoted passage, Kant describes the will as being "itself most highly legislating" (*selbst zuoberst gesetzgebend*, G 4:432). Furthermore, he repeatedly calls this legislation the will's "own (*eigene*) legislation" (G 4:431–2) and writes that it has "sprung from" the will (G 4:434). These and many other related expressions emphasize the *self* as the *source* of legislation, whereas Kant does not write anywhere in the *Groundwork* that the will "legislates to itself" in the sense that the self is the primary addressee of the laws.

Second, many similarly constructed expressions also indicate the *source* of a certain activity. The 'auto' in 'automobile' indicates that the vehicle is 'self-moving' in the sense that it moves by itself (as opposed to being moved by something else, such as a horse). An 'autocrat' rules by himself – and indeed, elsewhere Kant uses the adjective '*selbstherrschend* ' or 'self-ruling' as the German equivalent of the Greek-derived 'autocratic' (MdS 6:341). Here again the addition of 'self' emphasizes that the activity at issue (ruling) is done *by the self*, not that the self is the target of the activity.

Third, understanding 'autonomy' and 'self-legislating' in the sense of legislation *by oneself* yields a very straightforward contrast with 'heteronomy,' which, after all, indicates that one is subject to laws given *by another* (G 4:441). If autonomy is understood as legislation to oneself, this contrast is harder to construe.

Fourth, and relatedly, Kant in fact employs a *third* term to refer to legislation *to oneself*, namely 'heautonomy.' He uses this term in the *Critique of Judgment* to refer to the fact that the power of judgment "prescribes a law […] to itself" (*ihr selbst* […] *ein Gesetz vorschreibt*) (KU 5:185–6). Here we really do find legislation *to oneself*, and Kant calls it not 'autonomy,' but 'heautonomy.' Tellingly, he distinguishes 'heautonomy' from 'autonomy' by saying that the 'autonomy' of the power of judgment would involve the latter's prescribing a law *to nature* (*ibid*.). On the conception of autonomy as involving legislation *to oneself*, this passage is incomprehensible. It fits well with the account I propose, however, since, in parallel fashion, autonomy as a property of practical reason consists in its prescribing laws *to all rational beings*, and, similarly, the legislation analogy at the core of the Principle of Autonomy involves the idea of legislating *to all*. In sum, the German adjective *selbstgesetzgebend* in the passage at issue is not best read as meaning 'legislating *to* oneself.'

The common reading of autonomy is perhaps also due, in part, to the widespread assumption that Kant's conception of autonomy is best understood in terms of Rousseau's theory of political freedom, according to which freedom consists in living under the law one has prescribed

to oneself.[14] Many commentators view Kant's theory of autonomy as the application of this idea to the moral realm. They assume that Kant – in the mid 1780s – *agrees* with Rousseau. This brings us to the issue of Kant's position on the proper role of citizens in legislation. The core of the Principle of Autonomy is an analogy with political legislation: the Principle requires us to consider our will as simultaneously giving universal laws through our maxims. As I have explained above, this raises a question about consent. If we are to conceive of ourselves as moral legislators, are we to conceive of ourselves as simply imposing our maxims on others without their consent? Where does their consent enter into the picture, if at all? To get a clearer sense of Kant's conception of just legislation, we should examine the basis of the legislation analogy in Kant's political theory.[15] As it turns out, this is also crucial for understanding why Kant later dropped the Principle of Autonomy.

3 THE BASIS OF THE LEGISLATION ANALOGY IN KANT'S POLITICAL THEORY IN 1784

The Criterion of Just Legislation

The most extensive source for Kant's political theory around the time of the *Groundwork* and the *Critique of Practical Reason* is the set of student notes from his 1784 Lectures on Natural Law, known as the *Naturrecht Feyerabend*. Kant taught this course over the very months in which he wrote the *Groundwork*, which means that these lectures form a crucial resource for understanding the political analogies used in the latter. Moreover, the picture that emerges from the lecture notes is corroborated by published writings from the same year, such as Kant's 1784 essay "What Is Enlightenment?"

Kant agrees with the common notion – itself an ancient idea – that laws must be general (or 'universal') in scope and content. Regarding the criterion governing the *justice* of a law, Kant's position in 1784 is that a law

[14] Jean-Jacques Rousseau, *The Social Contract*, in *"The Social Contract" and Other Later Political Writings*, ed. Victor Gourevitch (Cambridge: Cambridge University Press, 1997 [1762]).

[15] My focus in this chapter is the analogy's basis in Kant's political theory. I do not mean to imply that this is the only relevant background for a full understanding of the emergence of Kant's account of autonomy. For example, developments in Kant's conception of law as such, and his conception of the parallels between laws of nature and moral laws, certainly also played a crucial role. See Eric Watkins, "Autonomy and the Legislation of Laws in the *Prolegomena*," in *The Emergence of Autonomy in Kant's Moral Theory*, ed. Stefano Bacin and Oliver Sensen (Cambridge: Cambridge University Press, 2018).

is just if it is *possible* for the people as a whole to impose it upon itself. He writes as follows:

> The touchstone of whatever can be decided upon as law for a people lies in the question: whether a people *could* impose such a law upon itself. (WiA 8:39, emphasis added)

In the *Naturrecht Feyerabend* lectures, he similarly argues that state laws are just if they *could* stem from the agreement of all:

> One must represent all laws in a civil society as given through the vote [*Stimmung*] of all. The original contract [*contractus originarius*] is an idea of the consent of all that has become a law for them. One must examine whether the law could have arisen from the agreement [*Uebereinstimmung*] of all: if so, then the law is right [*richtig*]. (NF 27:1382)

Importantly, Kant does not claim that just legislation requires citizens' actual consent: for a law to be just, it suffices that it "could have" arisen from general agreement or – in a slightly different formulation – that it be given "as if from general agreement" (EzA, 19:346). In other words, a political legislator can give laws that are fully just even if he does not have the actual consent of any subjects. The possibility of general consent suffices.

On Kant's view, the criterion of possibility – that is, the criterion governing *whether* a people could impose a law upon itself – is that the law is indeed truly general, or, in other words, that it is a genuine law. By implication, all genuine laws are just.[16] If a law is a genuine law – if it does indeed meet the generality criterion – then the people could give it to itself, and if the people could give it to itself, then the law is just.

To put it pointedly, this means that a law can be fully just even if it is laid down by a despot, if only the despot is enlightened enough to give laws that meet the criterion. Kant indeed reportedly claimed as much:

> The laws of a despot can be just, when they have been made such that they could have been made by the entire people. […] *It is not necessary* for him to judge whether the people *would* make such a law in this case, but whether it [viz., the people] *could have* made such a law. (NF 27:1382, emphasis added)

Thus, around the time that Kant wrote the *Groundwork*, he argued that the justice of a law is entirely independent of whether the people who are subject to it would actually choose to adopt it when given the opportunity.

[16] Rousseau draws a similar conclusion, although his argument for it differs, in Jean-Jacques Rousseau, *The Social Contract*, 2.6.

There is a direct fit between this criterion for the justice of political legislation and the *Groundwork*'s criterion for the moral permissibility of maxims. The Principle of Autonomy in its various formulations requires that one regard oneself as giving universal laws through one's maxims. It requires one to act as if one's maxims were at the same time to become laws for the entire moral community – that is, to act only on maxims that can simultaneously serve as universal laws. Kant can use the political criterion as the basis for formulating a criterion for the moral permissibility of maxims since both in the political realm and in its moral analog the decisive question concerns the required universality of the law. In neither the political realm nor its moral analog does the normative criterion require the actual consent of those who are subject to the law.

The Categorical Imperative as Constitution

All too often, authors assume that the law Kant describes as self-legislated is the Categorical Imperative. On closer inspection, however, it is clear that this cannot be correct.[17] For one thing, the Principle of Autonomy is itself a version of the Categorical Imperative. Furthermore, the Principle of Autonomy speaks of (as it were) *giving universal law through one's maxims*, or of considering the maxims of my will simultaneously as universal laws. Since the "universal law" mentioned in the Principle is the universalized version of my maxim, it cannot be the Categorical Imperative.

It helps greatly to keep in mind that Kant is using political analogies in his moral theory. I quoted Kant above as saying that states should be conceived on the model of an "original contract," by which he means that they should be conceived as having originated in the agreement of the subjects. Kant regards this idea as an a priori constitutional principle (NF 27:1382). He emphasizes that the idea of an original contract should not be misunderstood as a factual claim about the historical origin of states. Instead, it is a normative constitutional criterion for the justice of state legislation. At this stage of the development of his political theory, Kant regards the constitution as an a priori idea of reason, not as a matter of actual legislation, let alone a matter of actual consent by the citizens (NF 27:1382).

[17] Marcus Willaschek and I argue in defense of the thesis that Kant does not claim that the Moral Law or the Categorical Imperative is self-legislated, providing the necessary exegetical details in "Autonomy without Paradox: Kant on Self-Legislation and the Moral Law" (unpublished manuscript, 2017).

The role of the Categorical Imperative in moral legislation is parallel, then, to that of the state *constitution*.[18] Kant explicitly refers to the Categorical Imperative as the "constitutional law" (*Grundgesetz*) of pure practical reason (KpV 5:30) and as the "constitutional law" of an intelligible world (KpV 5:43). Just as the political constitution formulates a formal criterion for political legislation, the Principle of Autonomy – as one of the versions of the Categorical Imperative – articulates a formal criterion for the moral permissibility of one's maxims. At the time of writing the *Groundwork*, Kant regarded both the political constitutional law and the Categorical Imperative as a priori principles of reason.

4 KANT'S SECOND THOUGHTS ON CITIZEN CONSENT

The absence of 'autonomy' from the set of foundational concepts and principles in the *Metaphysics of Morals* does not seem to be a matter of debate in the Kant literature. In prominent recent volumes on the *Metaphysics of Morals*[19] and on Kant's conception of moral autonomy,[20] authors seem to proceed on the assumption that Kant is still committed to his previous theory of autonomy and simply fails to mention it.

Yet it is not at all evident that this assumption is correct. The number of occurrences of 'autonomy' drops off dramatically following the publication of the *Critique of Judgment*; thus, in fact, it disappears from center stage well before the *Metaphysics of Morals.* The term appears twenty-eight times in the *Groundwork*, fourteen times in the second *Critique*, and ten times in the *Critique of Judgment* (mostly in connection with taste). After this point, the term occurs sporadically in its political sense, for example when Kant mentions the autonomy of states (ZeF 8:346) or the autonomy of the university (SF 7:17). The word does not occur in discussions of moral theory such as those in the *Religion* and "On the Common Saying." As mentioned above, the term appears twice in the Tugendlehre of the *Metaphysics of Morals* – about which more below – but *not* to describe the "principle of morality" or the Principle of Autonomy. These facts shift the burden of proof to those who assume that Kant is still fully committed to his earlier theory, including the Principle of Autonomy.

[18] On this point, see also Andrews Reath, *Agency and Autonomy in Kant's Moral Theory*, 109–13.

[19] See, for example, *Kant's Metaphysics of Morals: A Critical Guide*, ed. Lara Denis (Cambridge: Cambridge University Press, 2010).

[20] See, for example, *Kant on Moral Autonomy*, ed. Oliver Sensen (Cambridge: Cambridge University Press, 2013).

It would be much more satisfying, however, to be able to *explain why* the Principle disappears. To this end, I will sketch the relevant changes Kant made to his political theory (that is, to the basis of the legislation analogy) before showing how these changes explain why Kant dropped the Principle of Autonomy.

Changes to the Basis of the Analogy

During the 1790s, most notably in *Toward Perpetual Peace* and the Doctrine of Right of the *Metaphysics of Morals*, Kant introduces the normative requirement that citizens actually consent to legislation. He drops the thesis that it *suffices* that the people *could* agree to a law and now adds the further requirement that the citizens also *do* agree to it (through their elected representatives in parliament). This is a major change to his theory, since in the *Naturrecht Feyerabend* he maintained that state laws do not require actual citizen consent in order to be just. I will not trace all the steps taken by Kant between 1784 and the 1797 *Metaphysics of Morals*, but I will highlight a number of important components of his later view for the sake of comparison.[21]

In the Doctrine of Right of the *Metaphysics of Morals*, Kant no longer holds that the laws of an enlightened despot can be fully just. He no longer regards it as sufficient that a law is truly general and *could* be adopted by the people as a whole. He now regards this as a merely necessary condition, adding the further requirement that legislation must result from *actual* consent by the citizens, via their elected representatives in parliament. Kant still attributes the status of an a priori normative principle to the "idea" of the "original contract" (MS 6:315), but he now adds that the "only constitution that accords with right" is that of a "pure republic" (MS 6:340). By this he means a "*representative* system of the people, in order to provide it with its rights, in its name, by all the state citizens united, by means of its delegates (deputies)" (MS 6:341). This ideal of the pure republic includes legislation by the citizenry, through their elected representatives.

[21] Elsewhere I discuss other aspects of Kant's political theory that underwent related changes around the same time, such as Kant's views on colonialism and the nature and approximate realization of an international federation. See Pauline Kleingeld, "Kant's Second Thoughts on Colonialism," in *Kant and Colonialism: Historical and Critical Perspectives*, ed. Katrin Flikschuh and Lea Ypi (Oxford: Oxford University Press, 2014), pp. 43–67, and Pauline Kleingeld, *Kant and Cosmopolitanism: The Philosophical Ideal of World Citizenship* (Cambridge: Cambridge University Press, 2012).

Kant writes that the legislative authority belongs to the united will of the people, that those members of a state who are "united for the purpose of legislation" are called "citizens," and that "the only qualification for being a citizen is being fit to vote" (MS 6:314). Kant infamously holds that fitness to vote depends not only on being an adult but also on being economically independent and male. He distinguishes between (propertied) "active" and (economically dependent) "passive" citizens, a distinction also found in the 1791 French constitution. He admits that this distinction seems to contradict his own account of citizenship and emphasizes that the law should leave it open to males to work their way up from passive to active citizenship (MS 6:314–15), but he rules out this option in the case of women.

Active citizens have the right to vote and elect their representatives to the legislature. In this way, they have the right to consent to legislation, albeit indirectly, by being "represented by [their] deputies (in parliament)" (MS 6:319).[22] The freedom of the citizen, Kant writes, is the attribute of "obeying no other law than that to which he *has* given his consent" (MS 6:314, emphasis added). Thus, he now argues that right requires that the citizens *do* consent (not individually but collectively, in parliament[23]) to the laws to which they are subject. Citizenship comes with more rights, of course, than the mere right to vote – for example, Kant also writes that active citizens have the right "to manage the state itself as active members of it, to organize it or to cooperate for introducing certain laws" (MS 6:315).

Kant is still committed to the view that there are a priori normative constraints on political legislation, because he still regards genuine universality as a formal normative requirement. Thus, compared with his position in 1784, the crucial difference for the purposes of this chapter is the addition of the requirement that the laws be given by the ('active') citizens themselves, through their elected representatives.

The Obsolescence of the Legislation Analogy

As a result of these changes, Kant's earlier analogy with political legislation became *unsuitable* as a formal model for assessing the moral

[22] The reason why Kant regards representation as necessary is that it makes possible the separation of powers (MS 6:341, ZeF 8:352).

[23] In practice, only a majority of votes is required, not unanimity; for Kant's defense of this principle, see TP 8:296.

permissibility of maxims. Using the old analogy would have been awkward. Kant came to *reject* the idea that legislation can be fully just without the actual consent of the citizens, and it would have been strange for him to articulate the principle of morality in terms of a political model he had discarded.

More importantly, the normative principles governing moral and political legislation are no longer fully analogous. The criterion of genuine universality remains in place in both domains. But whereas Kant now adds a new criterion for just legislation in the political domain, namely that laws be adopted by the citizens themselves (via their elected representatives), he does not add a parallel requirement in the moral domain. He continues to hold that maxims are morally permissible if they can simultaneously hold as genuinely universal laws. As a result, the normative constraints on political legislation are no longer suitable for articulating the moral constraints on one's maxims of action.

We can see this very clearly when we consider Kant's descriptions of the principle of morality in the *Metaphysics of Morals*. Here it is clear that he has not significantly changed the substance of the moral requirement. This is how he articulates the Categorical Imperative:

> Act on the basis of a maxim that can simultaneously hold as a universal law! (MS 6:225; cf. also 226)

According to this principle, it still suffices that one's maxim "can" simultaneously hold as universal law. Kant does not introduce an actual consent requirement into his moral theory.

Incidentally, in the *Religion* Kant explicitly addresses the disanalogy between the political and the moral realms. He writes that in the case of a republic or "juridical commonwealth," the people as a whole "is itself the legislator," but he denies that this is true in the case of the "universal republic according to laws of virtue" or the "ethical commonwealth." In the moral republic, he writes, "the people as such cannot itself be regarded as legislating"; the legislator must be "someone other than the people," namely God, who gives only "genuine duties" (Rel 6:98–9). Kant here seems to indicate that it is impossible to update the *Groundwork* idea of the "realm of ends" in the form of a moral analog of his new political ideal of a republic in which the united citizens themselves legislate.

In sum, Kant's political theory changed in a way that rendered the analogy with political legislation unsuitable. The old legislation analogy became obsolete.

Qualifying As a Universal Law

Kant continued to use political analogies in his moral theory, of course, including the idea of the moral community as a "republic of virtue" or an "ethical commonwealth," the idea of legislation as such, the idea of "ruling" oneself, and many others. The fact that he abandoned the legislation analogy that had been the core of the Principle of Autonomy does not imply that he no longer regarded the political and moral realms as in any way analogous. It therefore makes sense to ask how, if at all, we might redescribe the old analogy in terms of Kant's new political philosophy.

At what point in the ideal political process, as Kant conceives it in the *Metaphysics of Morals*, do citizens ask the question whether a principle *can* hold as a general law? Kant still holds that political laws must be genuinely general. Thus, in his new political theory there is still a point at which the old question is raised. This is the moment at which citizens ask themselves whether a legislative proposal is a real *candidate* for legislation – that is, whether a proposed law meets the formal requirement of being genuinely general.

The political counterpart of the question a moral agent should ask about any maxim he is considering, then, is the question a citizen should ask about any law he would like to propose. This is the question whether it *qualifies* for political legislation. If a law is directed at serving merely private interests, gives certain groups hereditary privileges, or lacks the required *generality* in other ways, then it is exposed as a pseudo-law that fails to meet the formal requirements governing legislation (e.g., TP 8:292–3).

It turns out that this is exactly how Kant reformulates the procedure for testing the moral permissibility of one's maxims. One should ask, he writes repeatedly, whether one's maxim simultaneously "*qualifies*" for general legislation (MS 6:225–6, 389, 393, 451, cf. 214). Kant still draws a parallel with the political realm, but the focus of the moral procedure is now on the question whether one's maxim *qualifies* as a universal law (while at the same time serving as one's maxim). We also find this terminology in Kant's political writings, for example where he remarks, with regard to political laws, that "*inclination* and, in general, what someone finds useful for his *private purpose* simply does not qualify as a law" (SF 7:32). In other words, the moral imperative no longer directs agents to act as if they were *giving* universal laws through their maxims; instead, it directs them to act on maxims that *could* become universal laws (and simultaneously serve as their own maxims).

It would be going too far to say that Kant here introduces an entirely new analogy. He does not elaborate the idea in any detail, perhaps because he wishes to avoid the suggestion that moral agents in the ethical community themselves legislate morally – the very suggestion he rejects in the *Religion*. Moreover, the idea that one's maxim should be "fit" or "appropriate" (*tauglich*, *schicklich*) for universal legislation is of course already implicit in the old legislation analogy in the *Groundwork* (4:438, 441, 444) and the *Critique of Practical Reason* (5:27–8, 36, 74). In the *Metaphysics of Morals*, however, it becomes the leading idea.

5 AUTONOMY IN THE *METAPHYSICS OF MORALS*

'Autonomy' occurs twice in the Doctrine of Virtue, both times without special emphasis, but in each case Kant speaks affirmatively of the autonomy of practical reason.[24] The task that remains is to show how these two passages fit with the narrative presented above.

The main argument of this chapter concerns the Principle of Autonomy, introduced in the *Groundwork* but absent in the *Metaphysics of Morals*. Neither of the two passages relates to the Principle of Autonomy as the formula of the Categorical Imperative that says that one ought to assess the moral permissibility of one's maxims by regarding oneself as simultaneously legislating universally through one's maxims. Instead, both passages concern autonomy as a *property*, specifically practical reason's property of being the source of moral laws independently of inclination. Kant is indeed still committed to the view that moral laws issue from reason itself. Hence, it is not surprising that he retains the idea of autonomy in this context.

The first passage is found in the introduction to the Doctrine of Virtue. Kant writes:

> For finite *holy* beings […] there is no doctrine of virtue but only a doctrine of morals, and the latter is an autonomy of practical reason, whereas the former also includes its *autocracy*, which is a […] consciousness of the *capacity* to master one's inclinations that rebel against the law. (MS 6:383)

Kant here calls a doctrine of morals "an autonomy of practical reason." He distinguishes between a "doctrine of morals" and a "doctrine of virtue," saying that the former but not the latter is applicable to *holy* rational beings. Such beings naturally do what morality requires and hence cannot

[24] There is a third occurrence of the term in the Doctrine of Right, where Kant refers to the autonomy of a state (MS 6:318).

be virtuous in the strict sense of the word. A doctrine of virtue does apply to humans, by contrast, because they are tempted to violate moral laws and need virtue in the sense of moral strength. This also means that a doctrine of virtue should include an account of how to master one's inclinations. The passage is not particularly easy to understand because the notion of "autocracy" remains somewhat ambiguous, but it is clear enough that the "autonomy of practical reason" refers to reason's property of being the source of moral legislation, independently of inclination. It does not refer to the Principle of Autonomy.

The second passage in which 'autonomy' occurs is found in the Doctrine of Method, in a discussion of why the imitation of good examples cannot ground a maxim of duty. Kant's answer is that a maxim of duty must be grounded in the "subjective autonomy of practical reason of each human being," which he explicates by saying that "the law must serve us as an incentive" (MS 6:480). Here he states that the mere imitation of others' outward behavior does not qualify as acting from duty; to qualify as an action from duty an action must be done on the basis of a maxim adopted from respect for the law. So what Kant seems to mean by "subjective autonomy" is a property of each agent's practical reason: namely, that it is the source of normative principles that provide incentives for moral action, independently of inclination. Again, this passage does not refer to the Principle of Autonomy.

In both passages, Kant seems to be claiming that practical reason has autonomy in the sense that it is the source of the (moral) laws to which rational beings are subject. From the *Groundwork* through the *Metaphysics of Morals*, Kant remains committed to this claim. The fact that Kant continues to use 'autonomy' in this context, even after he abandons the Principle of Autonomy, is neither strange nor problematic in light of the account provided above.

6 CONCLUSION

Around the time of the *Groundwork* and the *Critique of Practical Reason*, Kant claimed that genuine universality is a sufficient condition for a political law being just (regardless of whether the subjects would actually choose to adopt it if given the opportunity). During this period he regarded the criterion for the moral permissibility of maxims as fully analogous to the criterion for the justice of political laws. Accordingly, he formulated the Principle of Autonomy as requiring one to act as if, through one's maxims, one were simultaneously giving universal laws to the entire moral

community. During the 1790s, however, Kant changed his position on the criterion for just political legislation, adding the further condition that the citizens actually consent to the laws. As a result, the normative criterion for just political legislation could no longer serve as the basis for an analogy with which to express the principle of morality. This explains why Kant dropped the Principle of Autonomy.[25]

[25] I am grateful to Eric Watkins, Monique Hulshof, Carolyn Benson, Katharina Bauer, and Ken Westphal for helpful comments.