The Fate of Autonomy in Kant’s *Metaphysics of Morals*

Stefano Bacin (Università degli Studi di Milano)


**Abstract**: The idea of autonomy, presented as Kant’s main achievement in the *Groundwork* and the second *Critique*, is hardly present in the ethics of the “Doctrine of Virtue”. Against Pauline Kleingeld’s recent interpretation, I argue that this does not amount to a disappearance of the Principle of Autonomy, but to an important development of the notion of autonomy. I first show that Kant still advocated the Principle of Autonomy in the 1790s along with the thought of lawgiving through one’s maxims. I then argue that the role of autonomy in Kant’s later ethics has a different focus than in the previous works, which requires to connect autonomy with autocracy. Conversely, Kant’s ethics in the “Doctrine of Virtue” construes autocracy as a new layer of autonomy. Autonomy is there considered as autocracy, that is, as reason’s self-government in the maxims of each rational agent.

**Keywords**: Kant; autonomy; autocracy; self-government

1. A Principle Vanishes?

When Kant eventually published his long-awaited *Metaphysics of Morals*, the work that his contemporaries got to read did not exactly match what they had reason to expect after his previous works in practical philosophy. One of the numerous striking features of the new work is the lack of reference to the *Groundwork for the Metaphysics of Morals*, that is, to the work that had launched the project that the new work was supposed to bring to completion. Instead of the *Groundwork*, the first lines of the Introduction to the *Metaphysics of Morals*
mention the *Critique of Practical Reason* (VI 205). The *Groundwork* is never referred to in the entire work. An especially significant aspect of the conspicuous lack of connection to the previous work is the fact that the *Metaphysics of Morals* hardly devotes any space to one of its central upshots, namely the notion of the autonomy of the will. The very word ‘autonomy’ only occurs in two passages in the “Doctrine of Virtue”. One passage mentions autonomy in connection with the autocracy of practical reason (VI 383), while the other refers to “the subjective autonomy of each human being’s practical reason and so implies that the law itself [...] must serve as our incentive” (VI 480). After the emphasis put on the autonomy of the rational will in the *Groundwork* and the second *Critique*, the *Metaphysics of Morals* does not go any further than those two hints. Such a tenuous presence of the notion of autonomy in the final work of Kant’s practical philosophy is all the more striking considering that, in the years in which Kant worked on the *Metaphysics of Morals*, he was experimenting the possibility of applying the notion beyond the domain of practical philosophy, as is apparent in the drafts for the *opus postumum*.

---

1 The reference to the second *Critique*, furthermore, concerns not any specific claim of Kant’s moral theory, but the general need for a systematic exposition after a critique of reason: “The critique of practical reason was to be followed by a system, the metaphysics of *morals*” (VI 205). On this connection and the related issue see [reference removed]. — All references to Kant’s writings are given by volume and page number of the Academy Edition. For the *Ethik Kaehler*, I follow *Vorlesung zur Moralphilosophie*, edited by Werner Stark (Berlin–New York: De Gruyter, 2004). The English translation of the quotations is taken from the Cambridge Edition of the Work of Immanuel Kant, where available, and, for the *Groundwork for the Metaphysics of Morals*, from Jens Timmermann’s revision (Cambridge: Cambridge University Press, 2011).

2 A third passage, in the “Doctrine of Right”, concerns autonomy as property of a political community: see VI 318. Kant’s drafts for the *Metaphysics of Morals* also include only two mentions of autonomy: cf. XXIII 295 f.

3 In the *opus postumum* Kant mentions autonomy, for instance, as the key to the construction of physical reality (see XXII 78, XXII 398, XXII 404) or the possibility of sensible knowledge (see XXII 416). In several passages autonomy is even the central term in the definition of the nature and goals of transcendental philosophy (see e.g. X XI 59, X XI 79, X XI 93, X XI 106, X XI 108). On autonomy in the *opus postumum*, see Vaccarino Bremner, “Kant on the Autonomy of Reason”.

2
Although the path of Kant’s project in practical philosophy has been investigated from many different angles, the lack of a clear connection of the *Metaphysics of Morals* to the previous works with regard to autonomy has barely received attention. How autonomy featured in the post-Kantian debate has also been investigated, especially with regard to Fichte and Hegel. Yet, the fate of autonomy in Kant’s own later work has hardly been examined. The strikingly marginal role of autonomy in his later ethics has recently been pointed out by Pauline Kleingeld, who has also proposed an intriguing explanation for it. According to her, the notion of autonomy would experience a significant contraction after the second *Critique*, especially in the *Metaphysics of Morals*. In the *Groundwork*, autonomy is first introduced in a formulation of the categorical imperative that Kant calls the “Principle of Autonomy”: “act only so that the will could regard itself as simultaneously giving universal law through its maxim” (IV 434, cf. IV 432). Later on, Kant presents autonomy as the “the will’s property of being a law to itself” (IV 447; cf. IV 440). Now, in the *Metaphysics of Morals* a formulation corresponding to the Principle of Autonomy never occurs. Kleingeld argues that only autonomy as a property of the will retains a place in Kant’s later work, while the thought of “legislating through one’s maxims” that is conveyed in the Principle of Autonomy, “completely disappears”. According to Kleingeld, that is due to the fact that the Principle of Autonomy results from an analogy with political legislation. At the time of the *Groundwork*, Kant held that political laws can be just merely in virtue of their universal validity (cf. VIII 39; XXVII 1382). The possibility of general consent is sufficient to establish a genuine, according to Kant’s conception in the mid

---

4 One rare exception is Timmermann (“Duties to Oneself”, 214 fn. 19), who notes the marginal role of autonomy in the *Metaphysics of Morals*. However, he does not pursue the issue either. As Kleingeld (“The Principle of Autonomy”, 72) also notices, entire volumes on the *Metaphysics of Morals* do not mention this lack of an explicit connection, both to the autonomy of the will and to the *Groundwork* in general. See e.g. Denis, *Kant’s Metaphysics of Morals*; Engstrom, “The Inner Freedom”, 294-296.


1780s. As Kleingeld shows, Kant’s view would later experience a significant development, since he would later present a more demanding condition, arguing that political laws are in fact backed by the actual consent of the citizens (cf. VI 313 s.). In the *Metaphysics of Morals* and other writings of the mid-1790s, Kant argues that the citizens are subject only to laws to which they consent. Since Kleingeld holds that the Principle of Autonomy results from an analogy with the conception of political lawgiving, a change in that conception must impact on the analogy. Therefore, on Kleingeld’s interpretation, the analogy becomes obsolete: Kant could no longer coherently formulate the moral principle as the requirement that every rational subject consider oneself as a lawgiver for all rational subjects and the Principle of Autonomy accordingly disappears. Thus Kant articulates the principle only in terms of the qualification of maxims to a universal law (cf. VI 225 s., 389, 393, 451). In this interpretation, the change in Kant’s view on political legislation makes the Principle of Autonomy as we know it from the *Groundwork* vanish.

Kleingeld’s interpretation has the merit of highlighting a significant, yet hardly noticed issue in Kant’s use of the concept of autonomy. The alleged disappearance of the Principle of Autonomy should be re-examined, though. Here I shall suggest a different reconstruction of the fate of autonomy in the *Metaphysics of Morals*. For the textual evidence does not conclusively show that Kant discarded the thought of legislating through one’s maxims. Although Kant had reasons to use cautiously the Principle of Autonomy, he still advocated it. Moreover, assessing the place of autonomy in the “Doctrine of Virtue” requires taking the different task of that later work into account. A closer look at the relevant passages shows that the key to the role of autonomy in that work is its connection with the autocracy of reason, which in turn suggests a relative independence from Kant’s conception of political legislation. In Kant’s later ethics autonomy experiences not a contraction, but a development. As I shall show, autonomy is there considered not primarily as reason’s self-legislation, but as reason’s self-government in the maxims of each rational agent. The fate

---

of autonomy in Kant’s ethics consists thus in the emergence of his view of subjective autonomy.

2. Kant’s Cautious Commitment to the Principle of Autonomy in the 1790s

Since no passage in the *Metaphysics of Morals* alludes to the Principle of Autonomy, the fundamental requirement to consider one’s own maxims as legislating for all rational subjects appears to have been discarded. In Kleingeld’s interpretation, that is a result of the revision in Kant’s view of political laws. Before considering an alternate interpretation, however, we should ask if it is accurate to say that the Principle of Autonomy “completely disappears” in his later ethics.

Now, if he would abandon the Principle of Autonomy and the legislation analogy at its core, the very talk of lawgiving should have been entirely discarded as inadequate in ethics. Since lawgiving would require the actual consent by all beings who are subject to the relevant laws, Kant should not present something that does not satisfy that condition as lawgiving. Yet, Kant does present morality in terms of lawgiving throughout the *Metaphysics of Morals*. In fact, he introduces the general topic of the work by distinguishing two sorts of legislation (§ III, VI 218 ff.). Also, in the exposition of ethics, not only he refers to “our lawgiving reason” multiple times (VI 487; cf. VI 226, VI 406), but he applies the traditional juridical distinction between legislative and executive power to clarify the specific task of ethics (cf. VI 405). The prominence of the vocabulary of legislation seems hardly compatible with a full rejection of the thought of lawgiving through one’s maxims.

An accurate examination of the place autonomy has in Kant’s later ethics, however, must go beyond the *Metaphysics of Morals* and take also the lectures in

---

moral philosophy from the early 1790s into account. Those lectures include most of the mentions of autonomy after 1788, primarily because Kant presented to his students a survey of the main claims of the *Groundwork* and the second *Critique*. Kant reportedly explained that “the necessitation to duty results absolutely and unconditionally through the autonomy of reason” (XXVII 501). In that context, Kant also observed that “[m]oral legislation is the law-giving of human reason, as which it is the law-giver in regard to all laws, and is so through itself [*durch sich selbst*]” (XXVII 499). Thus Kant claims that “the moral laws must have their basic determination in a law-giving power which [...] constitutes legislation” (XXVII 499). Accordingly, he uses autonomy as the label for his own view and accordingly presents a “principle of autonomy” in contrast with the previous views on the foundations of morality (cf. XXVII 500).

Importantly, in his lectures from the 1790s Kant explains to his students that “the principle of the autonomy of reason” consists in “the very own [*selbsteigen*] legislation of choice by reason” (XXVII 499). This strongly suggests that the thought of legislating through one’s maxims had not been abandoned. Indeed, Kant advances the Principle of Autonomy twice in that course, formulating the categorical imperative as follows: “act so that you may present yourself, through the maxim of your action, as universally legislative” (XXVII 496; cf. XXVII 518). Notably, in the same course Kant puts forward his revised view of political legislation. For he explains that, “once it is established that man determines himself only through his own laws [...] and if it is only in such a fashion that we can understand how man may be bound”, we have to conclude that “the law [...] must be given with his own agreement, or at least must be capable of being seen as though it sprang from the united will of the

---

11 The so-called *Metaphysik der Sitten Vigilantius* goes back to a course held in 1793/94, at the time when Kant published the essay on *Theory and Practice*, one of the writings that Kleingeld (“The Principle of Autonomy”, 76) mentions with regard to Kant’s revised view on legislation.

12 Here I have modified Peter Heath’s translation for the Cambridge Edition, which has ‘individual’ for *selbsteigen*. *Selbsteigen* is merely intensifying for *eigen*, as the Grimm dictionary notes.
In spite of a revision to his view of political legislation, thus, Kant did not jettison the Principle of Autonomy. The 1790s lectures show that autonomy continues to include the content of that Principle, namely the thought of legislating through one’s maxims, which Kant still presents as the fundamental moral requirement.

This puts the role of autonomy in the *Metaphysics of Morals* in a different perspective. The merely negative evidence consisting in the absence of explicit mentions of the Principle of Autonomy in that work cannot let to conclude that Kant rejects it. That absence can be better explained if Kant’s caution in using it after the *Groundwork* is factored in. That caution was at least partially motivated by the observations of a critic. One of the important points made in 1786 by H.A. Pistorius, the reviewer who Kant appreciatively refers to in the second *Critique*, was that the Principle of Autonomy raised the suspicion of arbitrarism. Pistorius argued that, if Kant does not consider “to what extent a dissociation from all interest in willing from duty is indicated by this formula of autonomy, all this arbitrary [eigenmächtig] legislation seems a blind procedure [...], not that much different from what is usually called stubbornness [Eigensinn], of which one says: stat pro ratione voluntas”. Thus, Pistorius observed, the Principle of Autonomy seems to be at odds with the Formula of Universal Law, “for this formula does indeed [...] point at a condition that my maxim has to meet”. After that criticism Kant never used the Principle of Autonomy again without further clarification.

The contraction of the textual presence of the Principle of Autonomy results from the risk of a grave misunderstanding, to which even an insightful critic was prone. Since after Pistorius’ review, Kant always took special care in using the Principle of Autonomy, stressing its consistency with the Formula of Universal Law. This is precisely how Kant proceeds in the lectures of moral

---

13 Compare VI 318, VI 314, VI 315.
philosophy from the 1790s. There he states his principle combining both formulations: “act so that you may present yourself, through the maxim of your action, as universally legislative, i.e., so that the maxim of your action is suitable for universal legislation” (XXVII 496). Alternately, the Principle of Autonomy, presented as the main version of the “categorical general principle of morality”, is followed by the clauses that “morality can rest only on the law of reason” and can only concern “the form of lawfulness; for it must conform to the universality of the faculty of reason, and only under this form can it be morally good” (XXVII 518). Such passages witness to Kant’s care to prevent arbitraristic misunderstandings as well as his ongoing commitment to the Principle of Autonomy.

The pre-eminence of the Formula of Universal Law as the preferred way to phrase the categorical imperative in the “Doctrine of Virtue” must be considered against that backdrop. After Pistorius’s remarks, the second Critique had already contracted the presence of the Principle of Autonomy, well before any change in Kant’s view of political legislation. When Kant repeatedly mentions the fundamental requirement that maxims qualify for a universal legislation (cf. VI 225 s., VI 389, VI 393, VI 451), he follows the lead of the “fundamental law” of § 7 in the Critique, which is, after all, the work that the Metaphysics of Morals refers to in its opening lines. Kant’s carefully phrased, yet unambiguous endorsement of the Principle of Autonomy in the 1790s, thus, strongly suggest that the pre-eminence of the Formula of Universal Law in Kant’s later works does not entail that Kant was willing to drop the Principle of Autonomy.

Like Kleingeld, I stress that in the Metaphysics of Morals the principle of morality is always stated in terms of universal law, because I focus on the relationship between the Formula of Universal Law and the Principle of Autonomy, with the aim to clarify how Kant conceives of moral legislation. I do not mean to address thereby Allen Wood’s claim of the primacy of the Formula of Humanity in the derivation of ethical duties (see Wood, “The Final Form”, 12 f.). (Thanks to an anonymous reviewer for pressing me on this point.)

If Pistorius’s misunderstanding had a significant impact, the specific aim of the “Doctrine of Virtue” pushed Kant towards a different direction than a mere restatement of the Principle of Autonomy. In his treatment of ethics, Kant does not underscore the analogy with the main criteria of political legislation as in the *Groundwork*. Kant underscores that the general notion of legislation is not sufficient for the purposes of the ethical theory of the “Doctrine of Virtue”. At that point what *distinguishes* the ethical legislation from the juridical becomes relevant, as Kant immediately stresses in the introduction to the *Metaphysics of Morals* (§ III, VI 281 ff.). The emphasis on that distinction becomes only greater in the treatment of ethics.

The shift from the formulations of the previous works is already prominent in the drafts for the *Metaphysics of Morals*, in the contrast between a principle that imposes a limiting condition and a principle that gives a positive guidance to the will. Ethics needs the latter. Thus Kant notes: “That the maxim of my actions (subjective principle) be qualified to universal legislation is not the same as the principle that having that maxim is itself a duty. The former principle is merely limiting the faculty of choice [*Willkühr*], the latter is expansive [*erweiternd*]” (XXIII 392). In the “Doctrine of Virtue”, correspondingly, Kant remarks that the requirement that maxims “merely qualify for a giving of universal law [*allgemeine Gesetzgebung*] [...] is only a negative principle (not to come into conflict with a law as such)”. Kant argues that the task of ethics is to provide, “beyond this principle, a law for the maxims of actions” (VI 389; cf. VI 395). The universality requirement is the only formula of the categorical imperative that is explicitly mentioned in the “Doctrine of Virtue”, but only to be qualified. Kant emphasizes the distinctive feature of the ethical legislation, which does not set limits to actions, but prescribes specific contents for the maxims of every rational agent (cf. VI 388 f.; VI 382 f.).

The “law for maxims” that is the specific topic in the “Doctrine of Virtue” is legislated by each rational agent’s maxims. For Kant argues that ethics issues a
requirement that “is to be thought as the law of your own will and not of will in
general, which could also be the will of others” (VI 389). Ethics thus goes
beyond the mere demand of universal validity as it has to provide a positive
orientation through the content of one’s maxims. Accordingly, the specific
ethical principle stated in the “Doctrine of Virtue” sounds: “act in accordance
with a maxim of ends that it can be a universal law for everyone to have” (VI
395). By setting himself objective ends, a rational agent spells out how the
demands of morality are to be brought into effect in everyone’s maxims. The
ethical legislation not only addresses maxims, but is developed through
maxims, that is, through one’s maxims that are taken to enact the same contents
that every other rational agent should include in his own maxims. In this
respect, any virtuous maxim of an agent does state a law that every other agent
is expected to consent to. For, as Kant observes, imposing ends to other people
is impossible. They can only be embraced by one’s own will (cf. VI 381 f., VI
389). The law of ethics is thus focused on the necessity of actual consent in the
determinations of each rational agent.16

Kant’s view on the distinctive features of ethics suggests that ethics rules out
not the general conception of legislation employed in the Groundwork, but a too
narrow political analogy that would make the development of ethics strictly
dependent from views about juridical lawgiving. The development of Kant’s

16 Kleingeld claims that “Kant does not introduce an actual consent requirement into his moral
type” (Kleingeld, “The Principle of Autonomy”, 75), observing that Kant phrases the
principle of morality only as the requirement of universality, in the Metaphysics of Morals.
because she assumes that his revised view of political legislation, which requires actual
consent of the citizen for a law to be just, should have a direct impact on Kant’s treatment
of ethics. Note that, while Kleingeld claims that Kant’s statements of the principle of
morality in the Metaphysics of Morals do not include the requirement of actual consent,
she has later argued that Kant stresses the importance of actual consent in several passages
of the Metaphysics of Morals (see Kleingeld, “How to Use Someone ‘Merely as a
Means’”, 403). Thus that the “Doctrine of Virtue” does not phrase the principle of
morality in terms of actual consent, does not per se allow to conclude that at that point
Kant had given up the legislation analogy. On the contrary, the presence of the
requirement of actual consent in crucial places of the “Doctrine of Virtue” suggests that
the analogy is still very much in place and is only elaborated in accordance with the
specific aims of a treatment of ethics.
practical philosophy in the ethics of the “Doctrine of Virtue”, thus, does not leave the Principle of Autonomy behind, maintaining only autonomy as the property of the will. In that work, the thought of a moral legislation through one’s maxims, which Kant was still committed to, as the Vigilantius lectures show, is spelled out as the legislation of “a law for the maxims of actions” (VI 389; cf. VI 392 f.) through one’s will. The claim that ethics demands that the law is thought of as the law of one’s will in fact presupposes the thought that Kant first formulated in the Principle of Autonomy. Regarding oneself as a lawgiving member in the domain of morality imposes on each rational agent a fundamental constraint that leads to the positive guidance to give oneself maxims whose contents (ends) impose themselves on every other rational agent as obligations. The ethics of the “Doctrine of Virtue” thus unfolds a legislation that is given through the maxims of individual agents.

While the wording of the *Groundwork* emphasizes the role of the will in the universal legislation of the moral law, the “Doctrine of Virtue”, according to its different task, stresses that that legislation is to be enacted through one’s maxims. The underlying normative constraint is not phrased in terms of the Principle of Autonomy, but in terms of the Formula of the Universal Law, arguably also to avoid arbitraristic misunderstandings, as Kant already did prior to the *Metaphysics of Morals*. Still, if the lectures on moral philosophy show that the Principle of Autonomy as a formula of the categorical imperative is not abandoned in the 1790s, the way in which Kant spells out the task of ethics in the “Doctrine of Virtue” shows the thought of legislating through one’s maxims is not discarded either. More importantly, focusing on the task of ethics points us towards the right way to assess the role of autonomy in that final step of Kant’s project. The key is an *evolution* of autonomy, which in the “Doctrine of Virtue” is not merely retained, but takes on a more elaborate shape. I shall examine that important step in the following sections.
4. Autonomy in the “Doctrine of Virtue”: Subjective Autonomy and Autocracy

A differentiated account of the role of autonomy and legislation in Kant’s practical philosophy in the 1790s, however, must account for the fact that autonomy appears strikingly marginal in the “Doctrine of Virtue”. Kleingeld’s reading only focuses on the alleged disappearance of the Principle of Autonomy, without considering the positive role of autonomy in the new stage of Kant’s project. Yet, that role is what has primary importance and should be accounted for. The fate of autonomy in Kant’s later ethics should not be restricted to the limited role of the Principle of Autonomy, but should include in what capacity autonomy fits into the treatment of ethical duties. As I shall show, that fate does not merely amount to a sparse iteration of the notion of autonomy as a property of the will. In the “Doctrine of Virtue”, rather, Kant’s notion of autonomy adds a new layer to those explored in the *Groundwork*. Only two passages in the “Doctrine of Virtue” mention autonomy, as I have pointed out from the outset. The two passages are connected with each other because of an important common feature, which provides a crucial clue for the interpretation. Although in neither the Principle of Autonomy occurs, both passages do not merely reproduce the claims of the *Groundwork*, but introduce autonomy with regard to what Kant calls autocracy. If the first passage includes the only mention of autocracy in the “Doctrine of Virtue”, a careful reading of the second passage shows that autonomy is there used precisely to denote autocracy. I shall thus give a closer consideration to the two passages. In the first passage Kant mentions autonomy to underscore that his treatment of ethics cannot merely affirm the “autonomy of practical reason”, because it does address holy beings, but moral beings subject to temptation. Thus Kant characterizes a “doctrine of virtue” by explaining that it amounts to “autocracy of practical reason”, in contrast to mere autonomy. Since human beings as finite

---

17 On Kant’s view of autocracy in general see Baxley (“Autocracy and Autonomy”; *Kant’s Theory of Virtue*) and König (*Autonomie und Autokratie*).

18 This important connection between the two passages is overlooked by Kleingeld, who concludes that the only feature that they share is the absence of the Principle of Autonomy (see Kleingeld, “The Principle of Autonomy”, 77 f.).
rational agents face internal hindrances to the autonomy of pure practical reason and the law that it imposes, they have “consciousness of the capacity to master one’s inclinations when they rebel against the law” (VI 383; cf. XXIII 396). Kant had already observed many times that subjects affected by such limits cannot aim at holiness, but only at virtue, which entails an unavoidable conflict with an obstacle that can never be fully overcome. In the “Doctrine of Virtue”, this thought is eventually linked to a property of the will that is presented in close relation to autonomy. Autonomy is presupposed as the fundamental condition for rational agents to be subjects to moral obligations. Virtue requires that pure practical reason is also able to enact the obligating law given by reason in individual life, thereby overcoming the hindrances of inclination. As Kant writes, virtue requires not merely a legislative power, but also a corresponding executive power (cf. VI 405).

The connection with autocracy is the key to the role of autonomy in the “Doctrine of Virtue”. It is a further peculiar trait of the Metaphysics of Morals, however, that a notion that Kant introduces to characterize the task and the content of the work, that is, autocracy, never occurs again after that one passage. In the second passage in which autonomy is mentioned Kant uses the unexpected phrase ‘subjective autonomy’ instead, which, in turn, is not to be found anywhere else in his writings. In that second passage Kant claims that “a maxim of virtue consists precisely in the subjective autonomy of each human being’s practical reason and so implies that the law itself, not the conduct of other human beings, must serve as our incentive” (VI 480). The remarkable qualification of autonomy as subjective, here, shows both a continuity with the

19 See V 84. Kant has already made analogous claims more than twenty years before: see XXVII 13; XX 148, 151.
20 Baxley (Kant’s Theory of Virtue, 59) rightly stresses that the relationship between autonomy and autocracy is best interpreted via the distinction between legislative and executive power.
21 On the first passage Kleingeld (“The Principle of Autonomy”, 77) merely remarks: “The passage is not particularly easy to understand because the notion of ‘autocracy’ remains somewhat ambiguous”.

13
notion as it was introduced in the *Groundwork* and a novel development. By ‘subjective’ Kant denotes the individual dimension of the activity of reason in which moral requirements must be observed, that is, the determination of one’s maxims. A virtuous determination of the will cannot follow the example provided by the conduct of other people, but must be based on the observance of the law given by one’s own reason. The autonomy of the will requires that in every rational agent reason gives a law that is valid for all rational agents (cf. e.g. IV 438, V 36).

Now Kant argues that, because it bounds every finite rational agent, the law given by one’s own reason imposes itself on the individual rational subject as a normative principle that silences the hindrances to its requirement and directly shapes one’s maxims. While autonomy is the property of the will of being a law unto itself, the “subjective autonomy of each human being’s practical reason” is the capacity of pure practical reason of being an effective law to itself for each finite rational agent in her own maxims.

Kant thus describes in terms of autonomy what he understands as autocracy in the “Doctrine of Virtue”. In the second passage about autonomy he reiterates the wording in which, at time of the *Groundwork*, he had explained that “[w]hen reason determines the will through the moral law it has the force of an incentive, and then it has not merely autonomy but also autocracy”. In his moral philosophy lessons of 1785 Kant observed that autocracy is the property that reason enjoys when it has “both legislative and executive power” (XXIX 626; cf. Kaehler, 206; XXVII 360). The same terms come back in the “Doctrine of Virtue”, when Kant claims that in virtuous maxims reason is not merely lawgiving, but provides also the necessary enactment of its law against hindrances (VI 480). Thus, although autocracy is not explicitly mentioned there,

---

22 Oddly enough, when Dörflinger (“Ethische Methodenlehre”, 392) considers this remarkable passage, he does not consider the qualification of ‘subjective’. In fact, he does not even quote it, thereby missing a crucial element in Kant’s claim.

23 Subjective ”has here the same meaning as in the Doctrine of Method of the second Critique, which deals with “the way in which one can make objectively practical reason subjectively practical as well”, that is, “the way in which one can provide the laws of pure practical reason with access to the human mind and influence on its maxims” (V 151). See [reference removed]. For similar passages see V 38, 72, 74-76, 81, 88, 117.
it is what is thereby described in the second passage. If in the first passage about autonomy in the “Doctrine of Virtue”, autocracy is introduced as a further property of reason beyond mere autonomy, yet continuous with it, in the second passage the striking label ‘subjective autonomy’ stands for the same property.

The connection between autonomy and autocracy that is distinctive of the “Doctrine of Virtue” displays the limits of political analogies as a clue for the interpretation of Kant’s account of morality. As he did when he introduced the notion of autonomy in moral philosophy, Kant draws again on the political vocabulary when he connects it with autocracy. In fact, he had already employed the term ‘autocracy’ about morality not only before the “Doctrine of Virtue”, but even before the Groundwork. As several passages from his lectures on ethics show, Kant borrowed ‘autocracy’ from the political vocabulary already in the mid 1770s, before appropriating ‘autonomy’ as well while working at the Groundwork. In fact, the use of autocracy in morals might even have suggested to employ in morals another political term, autonomy. In the “Doctrine of Virtue” the connection with the political vocabulary is evident again when Kant observes that the new main focus is not a legislative power, but an executive power (cf. VI 383, VI 405; see also XXVII 499). However, this further connection with the political vocabulary shows that its use in the ethical domain does not merely follow the changes in Kant’s political views, contrary to Kleingeld’s interpretation. When Kant attributes autocracy to pure practical reason, he refers to a political form that he did not approve of. Kant explains in the “Doctrine of Right” that “the form of a state is either autocratic, aristocratic or democratic” (VI 338 f.; cf. VIII 352). Autocratic is, he argues, the appropriate name for what is usually called monarchic, because “an autocrat, who rules by himself has all the authority”. More precisely, “[t]he autocrat is the sovereign, whereas the monarch merely represents the sovereign.” Therefore Kant holds autocracy “with regard to right itself” the “most dangerous” form of a state “in

24 Here I find myself in agreement with Baxley (Kant’s Theory of Virtue, 59 f.), who argues that autonomy and autocracy cannot be regarded as a capacity and its realisation.
25 Kaehler, 206-212; XXVII 360-368; Refl 6867, XIX 186.
26 On the political background of the notion of autonomy, see Feil, Antithetik.
view of how conducive it is to despotism” (VI 339). Remarkably, that did not prevent him to describe his ethics in terms of the autocracy of practical reason. The asymmetry with Kant’s view about political autocracy significantly weakens the assumption that the moral notion of autonomy is dependent on his view on political legislation. Kleingeld argues that “it would have been strange for him [Kant] to articulate the principle of morality in terms of a political model he had discarded”. But articulating the requirements of morality in terms of an unsatisfying political model is exactly what Kant does in using ‘autocracy’ to characterize the main focus of the “Doctrine of Virtue” as such. Here, Kant does not rigidly follow an analogy, but applies the terms and roles of legislation to the ethical realm. The new connection between autonomy and autocracy rather results from a development of Kant’s own notion of autonomy in morality.

5. Autocracy as the Subjective Shape of Autonomy

However scarce, the presence of the term in the “Doctrine of Virtue” shows that Kant’s exploration of the notion of autonomy is not complete in the *Groundwork*. Since the task of the “Doctrine of Virtue” is obviously different from that of the *Groundwork* and the second *Critique*, the reader might well expect that autonomy does not play any role at all in the *Metaphysics of Morals*, since its conception only concern the foundations of moral obligation. Yet, the “Doctrine

---

27 Korsgaard (*Self-Constiution*, 153-157) refers to Kant’s discussion of the three forms of political government in the “Doctrine of Right” (VI 338-341), to argue that his conception of the state commits him to what she calls the Constitutional Model of the soul, according to which the person identifies not with reason, but with her constitution. However, Korsgaard does not point out that the political term that Kant eventually employs to describe the legitimate executive power of practical reason in ethics is autocracy, in contrast to his observations on the state. In spite of the limited correspondence of Kant’s treatment of ethics with his political conception, though, the role of the idea of autocracy in the “Doctrine of Virtue” does suggest a view close to Korsgaard’s Constitutional Model. The enactment of the moral law requires a capacity that does not consist in the mere force to overcome the hindrances of the sensibility, but lies in the power of practical reason to execute the moral law as a principle that silences those hindrances altogether.

of Virtue” does mention autonomy, and not merely in recapitulating claims of the previous works. As is appropriate to a novel idea, in fact, autonomy evolves after the *Groundwork* in the following steps of Kant’s moral philosophy, taking on layers that were not present in its previous treatments. The most important aspect of autonomy in Kant’s later ethics, thus, is not the mere permanence of the claim that the will is a law to itself, but how autonomy develops further.\footnote{On the contrary, Kleingeld only recognises in the “Doctrine of Virtue”, along with the alleged disappearance of the Principle of Autonomy, the mere reaffirmation of one claim from the *Groundwork*: “In both passages [in which autonomy is mentioned], 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” (Kleingeld, “The Principle of Autonomy”, 78).} What may appear as the fall of a principle in the “Doctrine of Virtue”, is in fact a new step in a longer progress. As I have shown, the Principle of Autonomy is not missing from Kant’s later ethics, and the contraction of its textual presence is primarily due to the need to prevent further misunderstandings. Most importantly, though, instead of evidence of the rejection of the Principle of Autonomy, the “Doctrine of Virtue” presents a development in Kant’s view of autonomy. Kant’s project in its final phase must focus not on reiterating the fundamental conditions of morality, but on clarifying how the principle of morality is to be enacted. Kant’s central claim is that this must happen not simply by applying a general principle to actions, but by making it “the law of one’s own will” (cf. VI 389), thereby effectively enacting the law given by reason through one’s maxims.

The fate of autonomy in Kant’s later ethics, thus, is determined by the connection with autocracy. Autocracy gives a further shape of autonomy as self-given legislation, which is now distinctive for this final step of his general project. Kant presents this thought in continuity with the general principle of morality. The proper task in a treatment of ethical obligations is to go beyond the fundamental moral criterion and provide possible moral maxims. Thus ethics builds on the principle of morality and “adds only that this principle is to be thought as the law of your own will and not of will in general, which could
also be the will of others” (VI 389; cf. VI 417). Autonomy and autocracy are not mentioned here, unlike a few pages before, in the first passage about autonomy (VI 383). Still, Kant’s claim is the same, and is stated in corresponding terms: Drawing on the self-given fundamental law of pure practical reason, ethics must present requirements that are valid for the subject insofar as they are regarded as articles of the law that the individual subject gives to herself. Autonomy of pure practical reason must be completed by its autocracy, in order to attain virtue.\(^{30}\)

When Kant remarks that his treatment of ethics should present the “autocracy of practical reason” (VI 383), the apparent contrast with autonomy points not at a separation between two properties, but at the necessity of completing one power with another.\(^{31}\) The power to enact a law is dependent on, and constrained by, the power to give a law to every rational agent. In turn, reason’s lawgiving is to be enacted in the maxims of each rational agent. Autonomy demands autocracy, which in turn is grounded on autonomy. Kant’s phrasing in the “Doctrine of Virtue” accordingly emphasizes the necessary connection between autonomy and autocracy, which is even closer than it might appear at first. Notably, whereas autocracy takes pride of place when Kant characterizes through it the overall task of the “Doctrine of Virtue” (cf. VI 383), it is not mentioned again in the rest of the work. The word ‘autocracy’ is also absent from the Vigilantius lectures. This contrasts with its recurrent presence in the earlier lectures. In fact Kant uses ‘autocracy’ mostly before 1785.\(^{32}\) After introducing autonomy as a key notion in his moral theory, Kant mentions

---

\(^{30}\) Note that the only other passage where autonomy and autocracy are mentioned together, in the 1785 Mrongovius II lectures, already follows the same pattern, presenting autocracy as a property of reason that goes beyond mere autonomy (cf. XXIX 626).

\(^{31}\) Guyer (“Kant on the Theory”, 143) has also maintained the continuity between autonomy and autocracy, observing that “[t]he achievement of autocracy […] is the only means that human beings have to implement the ideal of autonomy”. However, Guyer reaches this conclusion by referring mostly to the Collins lecture notes, which go back to lectures of the mid-1770s, when Kant had not yet worked out the notion of autonomy. As I shall clarify in a moment, it is necessary to distinguish the chronological and philosophical distance that separates the contexts in which Kant mentions autocracy.

\(^{32}\) Kaehler, 206-212; XXVII 360-368, XXVII 378, XXVII 392; Refl 6867, XIX 186.
autocracy hardly ever again in ethical contexts. Before introducing the conception of the autonomy of the will, autocracy simply means self-government or self-mastery, as Kant’s earlier lectures show. After the *Groundwork*, the pre-eminence and the fundamental role of autonomy made it necessary to connect the thought of autocracy to Kant’s novel view. In this light, autocracy primarily stands for an important issue that demands a solution, that is, how the moral law legislated by the autonomy of the rational will yields individual self-government. Kant’s final solution to that issue emerges in the “Doctrine of Virtue”, drawing on the notion of the autonomy of pure practical reason presented in the *Groundwork* and the second *Critique*.

The connection with the claim of the autonomy of the rational will gives the notion of autocracy a depth that it did not have before the *Groundwork*. Differently than in the earlier mentions of autocracy, thus, from 1785 on Kant examines the capacity to enact reason’s legislation in close connection to the corresponding legislative power, mostly without even mentioning autocracy. Kant refers to the executive power tied to the legislative power that had already been investigated in the previous steps of his practical philosophy, and even to “the subjective autonomy of pure practical reason” (VI 480), which is the role autonomy plays in enacting its own law in the maxims of an individual rational agent. Everything points at a continuity between the two roles of pure practical reason, which amounts to the emergence of a new layer of autonomy.

A further passage provides evidence in support of the continuity between autonomy and autocracy. In the unfinished essay on the *Progress of Metaphysics*

---

33 After the *Groundwork* and before the “Doctrine of Virtue” autocracy only occurs in the passage from Mrongovius II that I have quoted before (XXIX 626). A chronological differentiation has been absent so far from the studies devoted to Kant’s conception of the autocracy of practical reason. Baxley (*Kant’s Theory of Virtue*, “Virtue, Self-Mastery”) does not make any chronological distinction and brings together passages from rather different times in Kant’s work. Similarly, Guyer (“Kant on the Theory”) does not consider possible differences in the meaning of the notion of autocracy before and after the introduction of autonomy in 1785, nor the specific features of the notion of autonomy in the *Metaphysics of Morals*. 
Kant presents autocracy exactly as a further shape of autonomy. There he argues that reason can have “the power, in regard to its formal condition, namely morality, to attain this final purpose here in our earthly life, albeit as simultaneously intelligible beings, despite all the hindrances which the influence of Nature may exert upon us as sensory beings” (XX 295). This is precisely the property of reason that is presented as the main focus of the “Doctrine of Virtue”: the mastery of hindrances to which we are exposed as finite rational beings draws on the “formal condition” given by the moral law, which is thereby enacted in a specific domain, namely the maxims of each finite rational agent (cf. VI 383; XXIII 396). In that capacity, Kant writes, “the autonomy of pure practical reason is simultaneously taken to be autocracy” (XX 295). Autocracy is thus not a supplement to autonomy, but its necessary development in ethical life.

Autocracy thus constitutes the final shape of the idea of autonomy in Kant’s ethics, in which the normative commitment that is at the core of the notion of autonomy is made explicit and becomes the leading thought in the doctrine of ethical duties. In the Metaphysics of Morals autonomy as the property of the will of being a law to itself remains in place as the underlying assumption, but is now spelled out in the general task to enact that law as a law of one’s own will. The aim of ethics demands that autonomy as the property of the will determines the fundamental task for each individual agent. Autonomy demands, in the ethical life of rational subjects, autocracy, or “subjective autonomy”. This development brings the Groundwork’s claim of the autonomy of the will to imposes on finite rational agents a general requirement of self-government as the detachment from the tyranny of inclinations, which is only possible because of the property of the will of being a law to itself. The moral law legislated by the autonomy of the rational will can properly be brought to execution only by the same authority in the rational agent herself.

A further important aspect of the evolution of Kant’s conception of autonomy is, thus, that with the notion of autocracy, or subjective autonomy, Kant provides his own take on the issue of individual self-government. Kant’s notion
of the autonomy as presented in the *Groundwork* and the second *Critique*, does not correspond to what is currently called personal autonomy, since it does not concern the mere self-imposition of standards in individual choices.\(^{34}\) Now, the “Doctrine of Virtue” shows that, from Kant’s standpoint, what could be called personal autonomy, is in fact autocracy as the property of reason that makes it possible to enact its own law in the maxims of each finite rational agent. In contrast to current conceptions of personal autonomy, the distinctive element of Kant’s view is the continuity with the autonomy of the rational will. An individual rational agent has the authority needed for governing herself by virtue of reason’s moral legislation, as Kant’s analogy with the connection between legislative and executive power indicates (cf. VI 383 and XXVII 499): the latter can only be exercised on the ground of the norms given by the former.

Subjective autonomy, as the new, post-1785 shape of autocracy, shows that the traditional idea of self-mastery can only be incorporated into Kant’s view as the capacity of the rational will by virtue of which finite rational agents can set their ends and determine their maxims under the fundamental constraint of the moral law. The ethical self-mastery of individual rational agents, thus, is possible only as a development of the autonomy of reason.\(^{35}\)

6. Concluding Remarks

The suggestion that part of Kant’s notion of autonomy disappears because of a revision in his view of political legislation not only disregards his commitment to the Principle of Autonomy. More importantly, that suggestion overlooks the distinctive aspects of Kant’s project in the “Doctrine of Virtue”, namely the focus on a different legislation for maxims and the connection between


\(^{35}\) Note that, unlike personal autonomy in the current sense, Kant never presents autocracy as a property of the individual subject. As Baxley (*Kant’s Theory of Virtue*, 53) remarks, Kant ascribes autocracy to different subjects: practical reason, moral laws, freedom.
autonomy and autocracy. In spite of the political origin of that term, the analogy with political lawging proves not to determine the role of autonomy in Kant’s later ethics. Instead of the disappearance of an important part of Kant’s original conception, the scarce presence of autonomy in the “Doctrine of Virtue” rather displays a development of his idea. The thought of legislating through one’s maxims is in fact not given up, but evolves in the demand of enacting the law given by reason in one’s maxims. That the thought of legislating through one’s maxims still plays a role in the treatment of ethics in the “Doctrine of Virtue” shows that the conception of autonomy is considered by Kant not only as the key to the foundation issue of moral obligation, but also as the key to the enactment of the principle of morality by individual finite rational agents.

A closer examination of the relevant passages in the “Doctrine of Virtue” shows that the evolution of autonomy brings it to encompass also the executive power of the will that Kant otherwise calls autocracy. In this respect, the development of Kant’s view of autonomy amounts to an evolution of his notion of autocracy, which is thereby connected to the foundations of moral obligation. By focusing on the connection between autonomy and autocracy, the development of Kant’s conception of autonomy in the Metaphysics of Morals exhibits an important progress in three respects. First, the development shows that Kant’s thought of legislating through one’s maxims provides the key to understand how the individual rational agent should determine herself on the basis of the principle of morality. Second, it contributes to clarify the relationship between the cognate notions of autonomy and autocracy, which Kant had until then mostly employed in separate contexts. Third, the connection with autonomy and the idea of subjective autonomy sheds light on the notion of autocracy in Kant’s mature view, in spite of his limited elaboration of that aspect. Any complete account of Kant’s view of autonomy should thus include the layer added in the “Doctrine of Virtue”, that is, the thought of subjective autonomy as the role that reason assumes in enacting in one’s own maxims the law given by itself. Thereby Kant’s later ethics expounds his own version of what is currently
called personal autonomy. The *Metaphysics of Morals* shows that the fate of autonomy in Kant’s own work is to go through all levels of moral normativity.36

**Bibliography**


---

36 I would like to thank two anonymous reviewers for their helpful remarks.


