

Kant on Autonomy of the Will

Draft of a chapter to be included the *Routledge Handbook of Autonomy* (ed. Ben Colburn)

Biographical Note: Janis David Schaab is a Postdoctoral Researcher in Philosophy at the University of Groningen. He received his PhD from the Universities of St Andrews and Stirling in 2019. His research primarily focuses on Kantian approaches to ethical theory.

Abstract: Kant takes the idea of autonomy of the will to be his distinctive contribution to moral philosophy. However, this idea is more nuanced and complicated than one might think. In this chapter, I sketch the rough outlines of Kant's idea of autonomy of the will while also highlighting contentious exegetical issues that give rise to various possible interpretations. I tentatively defend four basic claims. First, autonomy primarily features in Kant's account of moral agency, as the condition of the possibility of moral obligation. Second, autonomy amounts to a metaphysical property as well as a normative principle and a psychological capacity. Third, although there is legitimate scholarly disagreement about whether or not autonomy involves self-legislation of the moral law, there is good reason to believe it underwrites an 'inside-out' (as opposed to 'outside-in') conception of the relationship between the will and moral requirements. Fourth, persons have dignity because their autonomy makes them members in the set of beings over whom the categorical imperative requires us to universalise our maxims, not because autonomy is an independently important property.

1. Introduction

In the *Groundwork of the Metaphysics of Morals*, Kant boldly asserts that all previous attempts “to detect the principle of morality [...] had to fail” for the same reason (G 4:432)¹:

One saw the human being bound to laws by his duty, but it did not occur to anyone that he is subject *only to his own* and yet *universal legislation*, and that he is only obligated to act in conformity with his own will which is, however, universally legislating according to its natural end.

According to Kant, therefore, the moral philosophers who came before him failed because they did not consider the possibility that the human being possesses *autonomy of the will*, which he defines as “the characteristic of the will by which it is a law to itself (independently of any characteristic of the objects of willing)”, and that the principle of morality is a “principle of autonomy” (G 4:440).

Since Kant thus takes the idea of autonomy of the will to be his distinctive contribution to moral philosophy, we might expect this idea, and the role it plays in his moral theory, to be clear and straightforward. This expectation might be reinforced by our confidence that we have a decent intuitive grasp of the concept of autonomy—which, after all, is central to much of current moral and political philosophy. However, Kant’s idea of autonomy of the will is more nuanced and complicated than one might think.

¹ I use the following abbreviations for (and translations of) Kant’s works: *Col* for *Collins’s Lecture Notes* (Kant 1997); *CPR* for *Critique of Pure Reason* (Kant 1999); *CPrR* for *Critique of Practical Reason* (Kant 1996); *G* for *Groundwork of the Metaphysics of Morals* (Kant 2011); *MM* for *Metaphysics of Morals* (Kant 1996); *Vigil* for *Vigilantius’s Lecture notes* (Kant 1997). I follow the pagination of the *Prussian Academy Edition*, except for the *Critique of Pure Reason*, where I follow the traditional A/B pagination.

In this chapter, I sketch the rough outlines of Kant's idea of autonomy of the will while also highlighting contentious exegetical issues that give rise to various possible interpretations. I will start by considering why Kant attributes such central importance to autonomy of the will (Section 2). On this basis, I try to gain a better understanding of what Kant means by 'autonomy of the will', particularly whether this term denotes a metaphysical property, a normative principle, a psychological capacity, or some combination of these (Section 3). Subsequently, I discuss whether and in what sense Kant's doctrine of autonomy should be understood to entail that we self-legislate the moral law (Section 4). Finally, I explore the relationship between autonomy and the moral standing of persons as ends in themselves (Section 5).

My focus throughout this chapter is on Kant's conception of autonomy of *the will*. I thus bracket Kant's remarks about the autonomy of *states*. Moreover, my goal is to sketch what autonomy of the will *means* for Kant and what role it plays in his foundational moral theory, as laid out, primarily, in the *Groundwork of the Metaphysics of Morals* and the *Critique of Practical Reason*. I do not discuss how Kant argues for the claim that we possess autonomy of the will, let alone how this claim relates to his overall metaphysical views. Finally, I do not consider how autonomy features in Kant's works on aesthetics, anthropology, or political philosophy. Any serious investigation of these issues would soon reveal that Kant's idea of autonomy is even more complex than I am able to show here.

2. Why Autonomy?

Why, according to Kant, does the fact that previous moral philosophers did not consider autonomy of the will as a possibility mean that their attempts to determine the principle of morality had to fail? The answer is that Kant holds that a "mere analysis of the concepts of

morality” (G 4:440) reveals that we must possess autonomy of the will if morality is to be more than a “chimerical idea without truth” (G 4:445). That is, the idea of autonomy is already entailed by ordinary people’s understanding of morality. In particular, autonomy is necessary to make sense of the idea of *moral obligation*, which is central to this understanding.

It is sometimes stated that the reason why Kant deems autonomy necessary for moral obligation is that he regards obligations as *unconditionally binding*. However, while this is basically correct, it is oversimplifying and potentially misleading. Without further clarification, the claim that moral obligations are unconditionally binding could be understood as the mere claim that they bind us regardless of our antecedent desires. But intuitionists like W.D. Ross, for example, agree with this claim and nevertheless deny that our will is a law to itself (Ross 2002: 157-160; see also Stratton-Lake 2013: 254-55). Ross holds that some acts are required or prohibited as a matter of “the fundamental nature of the universe” (ibid: 29). Thus, to understand why Kant attaches such great importance to autonomy, we need to determine *in what sense* he takes moral obligations to be unconditionally binding. And to determine this, we need to look at the entire bundle of moral concepts that he analyses in the *Groundwork*.

Kant’s analysis starts with the concept of the good will. He asserts that “[i]t is impossible to think of anything at all in the world, or indeed even beyond it, that could be taken to be good without limitation, except a GOOD WILL” (G 4:393). Kant grants that there are many other things that can be taken to be good, such as wit, confidence, and happiness. Yet, Kant argues, none of these things are good *unconditionally*. More specifically, they are not good unless they occur in the presence of a good will.

According to Kant, to unpack the concept of a good will, we should “inspect the concept of DUTY, which contains that of a good will, though under certain subjective limitations and

hindrances" (G 4:397). To imperfect rational beings like us, who are not solely determined by reason but also possess a sensuous nature, the actions that a good will would perform are represented as *duties*. More specifically, since our sensuous nature makes us susceptible to the influence of inclinations, the concept of a good will features as a *constraint* or "*necessitation*" in our deliberation (G 4:413). In short, it *obligates* us.

Kant argues that the limitations of imperfectly rational beings implicit in the concept of duty allow us to bring the concept of a good will into especially sharp relief (G 4:397). Imperfectly rational beings can act *either* from duty *or* from inclination. Thus, by *contrasting* the two sources of motivation available to them, we can get a better grasp of what it would mean to have a good will.

Kant argues that action "*from duty*" is to be distinguished, not only from action "contrary to duty", but also from action that is merely "*in conformity with duty*" (G 4:397-98). One of the examples that Kant uses to illustrate the latter distinction is that of an agent who helps others from sympathy, i.e. because their suffering pains him and its relief gratifies him. Kant asserts that, while this agent's action "conforms with duty", is even "amiable", it "has no true moral worth" (G 4:397). While the agent acts as duty requires, he does so only because, and insofar as, duty aligns with his inclinations. By contrast, the beneficent action of an agent who, through grief or by nature, is not inclined to act beneficently at all, has "far higher worth" (ibid.).

Kant takes this to show that, in imperfectly rational beings like us, a good will corresponds to the will of an agent who does his duty *from duty* alone, and thus *regardless of inclination*. He also takes it to show that "an action from duty has its moral worth *not in the purpose* that is

to be attained by it, but in the maxim according to which it is resolved upon, and thus [...] the *principle of willing* according to which [...] the action is done" (G 4:399-400).

Kant's reasoning seems to run as follows. The difference in moral worth between an action from duty and an action from inclination cannot be due to their respective purposes because they can have *the same* purpose (see Korsgaard 2008: 178; Timmermann 2007: 38). For example, both the sympathetic benefactor and the agent who acts beneficently from duty intend to relieve suffering.

Connectedly, for Kant, an action's purpose is problematically *external* to the agent's deliberation. In particular, an action's purpose or, more generally, an "object of the desiderative faculty" cannot determine the will *immediately* (G 4:400). If it determines the will at all, it does so *via* an inclination of the agent's. And we have already seen that, for Kant, actions are done *either* from inclination *or* from duty.

Note that the properties of the particular *act-type* that the agent performs are external to his deliberation in the same way as his purpose. Thus, they cannot be what determines the choice of a good-willed agent either. Accordingly, Kant remarks that moral worth does not reside in "the matter of the action" (G 4:416).

For Kant, then, actions from duty are guided, not by some external object or feature, but by something *internal* to the agent's deliberation. More specifically, they are guided by the quality of the agent's *maxim*. He states that "[a] *maxim* is the subjective principle of willing" (G 4:400, footnote). He later adds that a maxim is the principle an agent *actually* acts on, and thus to be distinguished from the *law*, which is the principle that the agent *ought* to conform to (G 4:420-21, footnote). What exactly Kant takes maxims to be is a difficult exegetical question (see Gressis 2010a; Gressis 2010b). What is important for our purposes is that Kant

thinks that the moral worth of an action depends on *which* maxim governs the agent's choice of the action.

But what characterises the maxim underlying a morally good action? Kant writes that "as every material principle has been taken away from it [...] nothing remains for the will that could determine it except, objectively, the *law* and, subjectively, *pure respect* for this practical law" (G 4:400). Kant's reasoning is that, since a morally worthy action is done from *duty*, and duty presupposes a *law*,² the maxim underlying such an action must be *lawful*. Therefore, an action from duty is an action from *respect for the law*.

However, the law in question cannot be a particular, substantive law. This is what Kant's statement that "every material principle has been taken away from [the will]" highlights (ibid.). It means that, in choosing how to act, a good will is not guided by the satisfaction of immediate inclination, the properties of a particular act-type, or the realisation of some purpose. Instead, the maxims underlying morally worthy actions are distinguished by their "mere conformity with law as such", i.e. their conformity to the very *idea* of law (G 4:402). In other words, what distinguishes these maxims is that they could *serve or function* as laws. We might say that they possess the *form* of laws (see G 4:436).

According to Kant, then, the lawful form of its maxims "alone is to serve the will as its principle, i.e. I ought never to proceed except in such a way *that I could also will that my maxim should become a universal law*" (G 4:402).³ Kant thus identifies the merely *formal* requirement only

² "The concept of duty stands in immediate relation to a *law* (even if I abstract from all ends, as the matter of the law)" (MM 6:389).

³ Why does Kant suddenly speak of *universal* law here? The reason seems to be that the law in question must be valid for *all* rational beings, independently of inclinations (see G 4:389).

to act on maxims that one could also will to become universal laws as the supreme principle of morality. For imperfectly rational beings, who are *necessitated* by the law, this principle takes the form of an *imperative*. Kant later identifies it as the only imperative that necessitates us independently of any inclinations and is therefore called “categorical” rather than “hypothetical” (G 4:420-21). He states the categorical imperative as follows: “*act only according to that maxim through which you can at the same time will that it become a universal law*” (G 4:421).

The categorical imperative leads us straight to the idea of autonomy of the will. Recall that Kant defines autonomy as “the characteristic of the will by which it is a law to itself (independently of any characteristic of the objects of willing)” (G 4:440). This is exactly the characteristic that the will must have if it is to be able to act from duty: it must select its maxims by a *law*, and yet independently of any particular, substantive law. To do so, it must select its maxims in light of whether they *could be willed* to be universal laws. Accordingly, Kant also refers to the categorical imperative as “the *principle* of every human will as a will *universally legislating through all its maxims*” (G 4:432) or simply as “the principle of autonomy” (G 4:440).

We can now see why it is so important to Kant that the will is regarded as autonomous. According to Kant, the concepts of duty and the good will, which are central to ordinary people’s understanding of morality, entail the concept of a law that binds *unconditionally*. Such a law is not merely independent of our antecedent desires (*pace* Ross), but necessitates us through pure respect for *lawfulness as such*. But this means that, if the concepts of duty and a good will are instantiated in the real world, we must possess autonomy of the will. Therefore, Kant’s charge that previous thinkers foreclosed the possibility of autonomy

amounts to the charge that they rendered *morality itself*, as understood by ordinary people, a mere *fiction*.

3. The Nature of Autonomy

Kant's idea of autonomy primarily characterises the way we, as *agents*, must relate to moral requirements if such requirements are to be possible. This, then, is the primary *function* that autonomy serves within Kant's framework. But what exactly does autonomy of the will *amount to*? Is it a metaphysical property, a normative principle, a psychological capacity, or some combination of these?

Kant argues that autonomy is both necessary and sufficient for *freedom of the will*. The will is free if it "can be efficient independently of alien causes determining it" (G 4:446). Thus, if the will can determine itself by its own lawful form, and thus independently of sensibility, it is free. Therefore, autonomy is sufficient for freedom of the will.

To show that autonomy is also necessary for freedom of the will, Kant argues that freedom has not only a "negative" but also a "positive" aspect (ibid.). According to Kant, "the will is a kind of causality" (ibid.). Now, "[s]ince the concept of causality carries with it that of *laws*" connecting causes with effects, this means that a free will cannot be "lawless" (ibid.). Accordingly, a free will is not merely *independent* of alien causes (and thus of the laws of *nature*) but proceeds in accordance with *its own law*. That is, it is autonomous.

As Karl Ameriks points out, Kant's notion of freedom of the will is "transcendental" rather than "empirical" (Ameriks 2013: 54-58). For the will to be independent from 'alien causes'—and thus to exhibit the negative aspect of freedom—it is not sufficient that it be independent from empirical causes *outside* ourselves, such as manipulation or external compulsion. Instead, the

will must not be determined by *any* empirical causes, including aspects of our empirical nature that we might ordinarily regard as expressing our individual nature or identity.

Thus, the association between autonomy and freedom of the will clearly shows that autonomy has a *metaphysical* dimension. But the association between autonomy and transcendental freedom also highlights another, *normative* dimension of autonomy (see, e.g., Reath 2013: 36; Sensen 2013b: 263). We can see this by shifting our attention to the side of autonomy that mirrors the *positive* aspect of freedom. Since autonomy is the characteristic of the will to be a *law* to itself, independently of the objects of volition, autonomy is not only necessary but sufficient for the authority of the moral law. Thus, Kant writes: “a free will and a will under moral laws are one and the same” (*G* 4:447). Henry Allison has labelled this the “reciprocity thesis” (1990: Chapter 11).

It follows that imperfect rational beings like us (who do not necessarily conform to the moral law), insofar as they are autonomous, are *bound* by the law, in the form of the categorical imperative. Thus, our will’s being a law to itself—its *ability* to be efficacious independently of empirical causes—has implications for what we *ought* to do. As Jens Timmermann puts it, “[o]ddly enough, ‘can’ implies ‘ought’” (2007: 121). We might even say that, for beings like us, autonomy itself *is* a normative principle. Indeed, Kant explicitly identifies “[t]he autonomy of the will as the supreme principle of morality” (*G* 4:440). For Kant, to say that the will is autonomous just *is* to say that it is subject to the “principle of autonomy”, i.e., “not to choose in any other way than that the maxims of one’s choice are also comprised as universal law in the same willing” (*ibid.*).

This highlights that autonomy is not only the condition of the possibility of moral obligation but also furnishes the *content* of moral commands. Since the moral law’s only prescription is

that our maxims be willable as universal laws, we might say that the moral law commands that we “realise” or “express” our autonomy (Sensen 2011: 169 and Hill 1992: 85, respectively). In Kant’s words, the law “commands neither more nor less than just this autonomy” (G 4:440).

But the positive aspect of freedom not only reveals the normative dimension of autonomy; it also implies that beings with autonomy must have certain *psychological* capacities. In particular, if a rational will is not only (negatively) independent of the laws of empirical nature but also (positively) a law to itself, rational agents must be capable of *recognising* and being *motivated* by that law. It thus seems that autonomy is not only a metaphysical property and a normative principle, but also “a recognitional *cum* motivational capacity” (Reath 2013: 33).

Note, however, that the precise psychological implications of autonomy can vary between different kinds of rational beings. For imperfect rational beings like us, who are sensuously affected, the capacity to be motivated by the law requires the existence of an *incentive* to comply with it (CPrR 5:72). Since this incentive must be able to counteract impulse and inclination, which are “based on feeling”, the incentive must itself be (based on) a feeling (ibid.). Yet, since this feeling could not enable autonomy if it was empirically conditioned, it must instead be “solely produced by reason” (CPrR 5:76). Kant identifies that feeling as “respect for the moral law” (ibid.).

By contrast, perfect rational beings (such as “the divine will”) do not require any incentive to be motivated by the moral law (CPrR 5:72). Since they are not sensuously affected, their actions necessarily conform to the moral law. Consequently, autonomy should not be *identified* with the capacity to act from respect for the law, although it *requires* that capacity in human beings.

4. Autonomy as Self-Legislation of the Moral Law?

So far, my focus has been on the autonomous will's ability to be *determined* by a law that is not empirically conditioned. But 'autonomy' literally means *self-legislation*. As Kant himself states, "the will is not just subject to the law, but subject in such a way that it must also be viewed as self-legislating, and just on account of this as subject to the law" (*G* 4:431). Hence, autonomy of the will seems to imply that the will in some sense *gives* law to itself.

This dimension of autonomy seems important because, as Andrews Reath observes, it suggests that autonomy underwrites "a normative standing of the rational will" (2013: 33). The autonomous will seems to be "normatively self-governing" in that "it need not answer to any outside authority, but only to itself" (*ibid.*) In short, autonomy underwrites a kind of "sovereignty of the will over itself" (Reath 2013: 49).

However, the idea that the will self-legislates the moral law, i.e. the law expressed by the categorical imperative, can seem puzzling. The puzzle is how to square this idea with the *unconditional authority* of the moral law. For talk of 'self-legislation' may seem to suggest that the law's authority is conditional on an act of legislation (Kleingeld and Willaschek 2019: 3; Pinkard 2002: 59; Stern 2012: 13-14).

This raises the question of how the 'legislation' part of 'self-legislation' is to be understood. Kant distinguishes two roles in legislative processes. He calls the one who *commands* a law its "lawgiver" (*MM* 6:227). Since the lawgiver makes the law *binding*, he is the "author of the obligation" (*ibid.*). But the lawgiver need not also be the "author of the law", i.e. the one who determines its *content* (*ibid.*) And, indeed, Kant seems to suggest that the moral law has *no author* (*Col* 27: 282–83; *Vigil* 27:544; see also, e.g., Kain 2004; Stern 2010; Timmermann 2007:

107; Wood 2008: 113). It is thus not up to us, nor anyone else, what the law commands. Instead, the content of the law is given by reason alone.

However, this qualification may not dispel the air of paradox surrounding the notion of autonomy completely. After all, it leaves open the possibility that it is up to us—the very agents who are subject to the law—to make the law *binding*. This leads us to a second question: who or what does the ‘self’ in ‘self-legislation’ refer to?

One option is that it refers to our self as concrete, empirical human beings. This interpretation might be congenial to theorists who associate autonomy with the expression of one’s personal values or identity (e.g. Bratman 1979; Frankfurt 1988). And, indeed, some contemporary Kantians are sometimes read as endorsing it (e.g. Korsgaard 2009). But this interpretation also makes it harder to see how the moral law can possess unconditional authority. For it suggests that the self that is subject to the moral law is indeed identical to the self that legislates it. Kant seems to rule out this interpretation for this very reason (*MM* 6:417).

It seems more likely that, instead, the ‘self’ refers to our noumenal self (Sensen 2013b: 268; Timmerman 2007: 115). While our phenomenal self corresponds to our nature as a concrete human being, our noumenal self corresponds to our nature as a being with pure practical reason, an immutable nature that we share with all other rational beings (*MM* 6:418). A related (though not identical) distinction in Kant’s works is between our faculty of “*choice*” (*Willkür*) and our “*will*” (*Wille*): while the former enables us to act and realise the objects of our desires in general, the latter enables us to be guided by pure reason alone (*MM* 6:213; also see Sensen 2013b: 269; Timmermann 2007: 115, 173-75). Thus, if our noumenal self or will—as opposed to our phenomenal self or faculty of choice—gives the law to us, this might

explain how we can be both self-legislating and subject to the unconditional authority of the law.⁴

A third option is that the 'self' in 'self-legislation' does not refer to one of the *agent's* selves at all. Instead, it indicates that the law is authoritative *of itself*, i.e. without presupposing anything else (Sensen 2013b: 269; also see O'Neill 2003: 15-19). The idea is that the moral law states the conditions of the possibility of lawgiving, and therefore does *not* depend on the authority of any *legislator*.

This interpretation suggests that the moral law is not 'legislated' in any *literal* or *straightforward* sense. But it does not abandon the assumption that the moral law is legislated in *some* sense, which is shared by a wide range of scholars (e.g. Engstrom 2009: 149; Reath 2006; Reath 2013; Schneewind 1998: 6; Wood 2008: 106). An outright rejection of this assumption has been suggested more recently, by Pauline Kleingeld and Marcus Willaschek (2019).

Kleingeld and Willaschek raise a third question about talk of 'self-legislation': *what*, if anything, is being legislated? They argue that Kant does not regard *the* moral law, as expressed by the categorical imperative, as self-legislated at all (2019: 2). He unequivocally speaks of self-legislation only in the context of *substantive moral laws*, i.e. imperatives that express the *specific duties* that result from *applying* the categorical imperative to our maxims. According to Kleingeld and Willaschek, Kant calls the categorical imperative "the principle of autonomy" because it commands that we select our maxims *as if* we were thereby also giving law to all rational beings (2019: 6-7). They conclude that Kant's notion of autonomy does not imply that

⁴ For a qualified criticism of this idea, see Reath (2013: 45-47).

the authority of *the* moral law is grounded in any kind of legislation. Instead, this authority is a fundamental, *a priori* fact of reason (2019: 13-16).

The latter three interpretations of the notion of self-legislation—that the *noumenal* self or *Wille* legislates the law, that the law is authoritative *of itself*, and that only *particular* moral laws are to be regarded as self-legislated—all seem like promising candidates when it comes to salvaging the *unconditional authority* of the moral law. However, these interpretations may not seem congenial to the idea with which this section started: that autonomy underwrites a kind of *sovereignty* of the will over itself.

Yet, one idea, which is popular among Kantian ethicists and seems largely compatible with these three interpretations, might help to alleviate this concern. The idea is that the moral law is the *constitutive* principle of willing or, equivalently, of the activity of practical reason. Different versions of this idea have been proposed in the literature (see, e.g., Engstrom 2009; Hill 1992: 87-88; Korsgaard 2009; Reath 2013; Sensen 2019).⁵ Here, I merely try to sketch what I take to be their unifying theme.

As Christine Korsgaard notes, to say that a principle is constitutive of an activity is to say that “unless you are guided by the principle in question, you are not performing that activity at all” (2009: 29). For example, the constitutive principle of building a house requires that you erect a physical structure that provides shelter from the weather (*ibid.*). If you do not at least aspire or strive to conform to this principle, you are not engaged in house-building at all.

⁵ One question that divides theorists is whether the law is constitutive of *Wille* (Sensen 2019), *Wille* and *Willkür* combined (Reath 2013), or *all* types of practical reason (Korsgaard 2009).

The constitutive principle of an activity is a *normative standard* for that activity (Korsgaard 2009: 29-32). For example, building a physical structure that has random holes in the walls and ceiling will at best count as building a house *badly* because it at most imperfectly conforms to the constitutive principle of that activity. However, whereas normative standards that are not constitutive of house-building (e.g. standards of beauty) depend on the presence of certain external factors (e.g. whether the house-builder cares about beauty), the constitutive principle of an activity seems to provide a normative standard which is *internal* to that activity.

Thus, the claim that the moral law is the constitutive principle of willing amounts to the claim that, unless we acknowledge the categorical imperative as binding, we are not engaged in the activity of willing at all. What is more, it implies that the categorical imperative provides a normative standard for our actions which is *internal* to our will. Far from imposing an *external* constraint on us, it expresses our own will's *in-built* ambition to be a law to itself.⁶

It is not implausible to attribute this view to Kant. As we saw above, Kant thinks that if our will possesses the *metaphysical* property of transcendental freedom (and the corresponding psychological capacity), we are bound by the *normative* requirement to make our maxims willable as universal laws. The idea that the categorical imperative is *constitutive* of the activity of a free will can help us make sense of this.

If this is Kant's view, then perhaps the idea that autonomy accords the will a kind of sovereignty can be salvaged after all. Even if we do not legislate the moral law to ourselves, it is in a sense *our own* law. We strive to comply with this law by virtue of being the kind of

⁶ This need not make *immoral* willing impossible. Agents who don't comply with the moral law might still be said to *acknowledge* its bindingness. Kant writes: "even the most hardened scoundrel [...] is conscious of [...] the law, the repute of which he recognizes as he transgresses it" (G 4:454-55).

creatures who engage in willing (see Reath 2013). Indeed, since it is an internal standard of willing, the law would not bind anyone if there were no creatures like us (see Sensen 2013c; Sensen 2019).

This suggests that Kant's idea of autonomy of the will parallels his so-called Copernican Revolution in theoretical philosophy (see *CPR* BXvi). As Oliver Sensen puts it, just as Kant's theoretical philosophy proposes that "we are not merely passive observers in cognizing nature, but that our mind (partly) constitutes how reality appears to us", his practical philosophy proposes that "it is an activity of reason, something reason does spontaneously and out of itself (not something that it discovers) that grounds morality" (2019: 199 and 217, respectively). We might say that Kant's idea of autonomy portrays the relationship between the will and moral requirements as 'inside-out' rather than 'outside-in'.

5. Autonomy and Dignity

I have argued that autonomy primarily features in Kant's theory of moral *agency*. But Kant famously claims that rational beings are "*persons*" rather than "*things*", "because their nature already marks them out as ends in themselves, i.e. as something that may not be used merely as a means" (*G* 4:428). And he writes that "[*a*]utonomy is [...] the ground of the dignity of a human and of every rational nature" (*G* 4:436). This suggests that autonomy also plays an important role in Kant's account of moral *patienthood*.

Furthermore, since Kant describes "*dignity*" as "an inner worth", contrasting it with "a price", which is merely "a relative worth" (*G* 4:435), it might seem as if the property of autonomy has *value* or some other kind of moral importance, independently of the moral law. And this might in turn seem to *explain* the special moral standing of persons. Some Kant scholars seem to come close to endorsing this interpretation (e.g. Caranti 2017; Guyer 1998). It might appeal

to theorists who want to derive human dignity or human rights from autonomy *directly*, without taking a detour through the categorical imperative (e.g. Griffin 2008). But this interpretation seems to contradict my claim that Kant proposes an ‘inside-out’ rather than an ‘outside-in’ relationship between rational agents and moral requirements.

However, there is good reason to doubt that Kant conceives of autonomy as an independently important property. As we saw in Section 2, Kant thinks that a morally worthy action cannot be one that is done for the sake of some “object of the desiderative faculty”, i.e. some object that lies outside the will and that the will is aimed at, because such an object can only determine the will *via* the agent’s inclinations (*G* 4:400; also see *CPrR* 5:62-64). Connectedly, actions done for the sake of such an object would not realise our autonomy. After all, when we act from inclination, we conform to the laws of empirical nature.

These considerations also apply to the notion of autonomy as an independently important property (Sensen 2013b: 276-79).⁷ Any action for the sake of such a property would be from inclination, and hence not realise autonomy. Kant holds that autonomy, and the ordinary conception of morality that relies on it, are possible only if the will has its own, *a priori* law. All objects of the will—even the property of autonomy itself—are to be assessed by that law.

But how is the moral importance of autonomy to be understood, if not in terms of an independently important property? Recall that the moral law requires that we act on maxims that we could will to be *universal* laws, i.e. laws that determine the will of every rational being. Hence, our will “must always take its maxims from the point of view of itself, but also at the same time of every other rational being as legislating (which are therefore also called

⁷ Also see Chapter 3 in this volume.

persons)” (G 4:438). In other words, the will’s maxims must be such that they could be willed as universal laws by—and thus realise the *autonomy* of—every rational being. And this, arguably, *just is* the requirement not to use others as mere means and instead to regard them as ends-in-themselves. The following passage seems to confirm this reading (CPrR 5:87, italics altered):

[E]very will [...] is restricted to the *condition of agreement with the autonomy of the rational being*, that is to say, such a being is not to be subjected to any purpose that is not possible in accordance with a law that could arise from the will of the affected subject himself; hence this subject is to be used *never merely as a means but as at the same time an end*.

The idea seems to be roughly that, in order to avoid using someone as a mere means, we have to act on a maxim that they could also rationally will (O’Neill 1989: 141-44; Reath 2013: 51-52; Sensen 2011: 109).⁸ And this is just what it takes for our maxim to be compatible with the categorical imperative.

Therefore, persons have a special importance or “worth”⁹, not because autonomy is an independently valuable or otherwise important property, but because the law commands that our maxims be such that *we* could will them to become universal law, which is the case if and only if they could be willed by *any* rational being as such. The dignity of persons is therefore grounded in their autonomy in the sense that their autonomy makes them members in the set of beings over whom our maxims must be universalisable (see Bader forthcoming). In

⁸ For a different reading, see Kleingeld (2020). For discussion, see Chapter 3 in this volume.

⁹ I bracket the question of whether the moral standing of persons should be framed in terms of *value* at all (see Bader forthcoming).

Kant's words, autonomy makes each person a "legislating member of the universal kingdom of ends" (G 4:438).

This corroborates my claim that autonomy primarily characterises the way persons, considered as moral *agents*, relate to moral requirements—namely as (fellow) legislators rather than as mere subjects. The moral standing of persons, considered as moral *patients*, merely reflects this relation. Hence, Kant writes, "morality and humanity, *in so far as it is capable of morality*, is that which alone has dignity" (G 4:435, italics added).

6. Conclusion

I have argued that autonomy primarily features in Kant's account of moral agency, as the condition of the possibility of moral obligation. Autonomy is at once a metaphysical property of the will, a normative principle (for imperfectly rational beings) and a psychological capacity. While there is reasonable disagreement about whether autonomy implies that the moral law is self-legislated, there is good reason to think that it underwrites an 'inside-out' picture of the relation between moral agents and moral requirements. These claims are not undermined by autonomy's relation to dignity. Persons have dignity because their autonomy makes them members in the set of beings over whom the law requires us to universalise, not because autonomy is an independently important property.

Further Reading

Oliver Sensen (ed.), *Kant on Moral Autonomy* (Cambridge: Cambridge University Press, 2013) is a useful collection of recent essays on Kant's conception of autonomy, its historical significance, and its relevance today. Stefano Bacin and Oliver Sensen (eds.), *The Emergence of Autonomy in Kant's Moral Philosophy* (Cambridge: Cambridge University Press, 2019) is a

collection of cutting-edge essays on the development of the idea of autonomy across the various stages of Kant's ethical thought.

References

Allison, Henry. 1990. *Kant's Theory of Freedom*, Cambridge: Cambridge University Press.

Ameriks, Karl. 2013. "Vindicating Autonomy," in Sensen 2013a: 53-70.

Bader, Ralf. Forthcoming. "The Dignity of Humanity," in Sarah Buss and Nandi Theunissen (eds.), *Rethinking the Value of Humanity*, Oxford: Oxford University Press.

Bratman, Michael. 1979. "Practical Reasoning and Weakness of the Will." *Noûs*, 13: 131–51.

Caranti, Luigi. 2017. *Kant's Political Legacy: Human Rights, Peace and Progress*, Cardiff: University of Wales Press.

Engstrom, Stephen. 2009. *The Form of Practical Knowledge*, Cambridge, MA: Harvard University Press.

Frankfurt, Harry. 1988. *The Importance of What We Care About*, Cambridge: Cambridge University Press.

Gressis, Robert. 2010a. "Recent Work on Kantian Maxims I: Established Approaches." *Philosophy Compass* 5(3): 216–27.

— 2010b. "Recent Work on Kantian Maxims II." *Philosophy Compass* 5(3): 228–239.

Griffin, James. 2008. *On Human Rights*, Oxford: Oxford University Press.

Guyer, Paul. 1998. "The Value of Reason and the Value of Freedom." *Ethics* 109: 22–35.

Hill, Thomas Jr. 1992. "The Kantian Conception of Autonomy," in *Dignity and Practical Reason in Kant's Moral Theory*, Ithaca and London: Cornell University Press: 76-96.

Kain, Patrick. 2004. "Self-Legislation in Kant's Moral Philosophy." *Archiv für Geschichte der Philosophie* 86: 267-306.

Kant, Immanuel. 1996. *Practical Philosophy*, Mary Gregor (ed. and trans.), Cambridge:

- Cambridge University Press.
- 1997. *Lectures on Ethics*, Peter Heath (ed. and trans.) and Jerome B. Schneewind (ed.), Cambridge: Cambridge University Press.
 - 1999. *Critique of Pure Reason*, Paul Guyer and Allen W. Wood (eds. and trans.), Cambridge, UK: Cambridge University Press.
 - 2011. *Groundwork of the Metaphysics of Morals: A German-English Edition*, Mary Gregor and Jens Timmermann (eds. and trans.), Cambridge: Cambridge University Press.
- Kleingeld, Pauline. 2020. "How to Use Someone 'Merely as a Means'." *Kantian Review* 25(3): 389-414.
- Kleingeld, Pauline and Willaschek, Marcus. 2019. "Autonomy Without Paradox: Kant, Self-Legislation and the Moral Law." *Philosophers' Imprint* 19(6): 1-18.
- Korsgaard, Christine. 2008. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, Oxford: Oxford University Press.
- 2009. *Self-Constitution: Agency, Identity, and Integrity*, Oxford: Oxford University Press.
- O'Neill, Onora. 1989. *Constructions of Reason: Explorations of Kant's Practical Philosophy*, Cambridge, UK: Cambridge University Press.
- 2003. "Autonomy: The Emperor's New Clothes." *Proceedings of the Aristotelian Society*, Supplement, 77: 1-21.
- Pinkard, Terry. 2002. *German Philosophy 1760–1860: The Legacy of Idealism*, Cambridge: Cambridge University Press.
- Reath, Andrews. 2006. *Agency & Autonomy in Kant's Moral Theory*, Oxford: Oxford University Press.

- 2013. “Kant’s Conception of Autonomy of the Will,” in Sensen 2013a: 32-52.
- Ross, William D. 2002. *The Right and the Good*, Philip Stratton-Lake (ed.), Oxford: Clarendon Press.
- Schneewind, Jerome B. 1998. *The Invention of Autonomy: A History of Modern Moral Philosophy*, Cambridge: Cambridge University Press.
- Sensen, Oliver. 2011. *Kant on Human Dignity*, Berlin/Boston: De Gruyter.
- 2013a (ed.). *Kant on Moral Autonomy*, Cambridge: Cambridge University Press.
 - 2013b. “The Moral Importance of Autonomy,” in Sensen 2013a: 262-81.
 - 2013c. “Kant’s Constructivism”, in Carla Bagnoli (ed.), *Constructivism in Ethics*, Cambridge, UK: Cambridge University Press: 63-81.
 - 2019. “Kant’s Constitutivism,” in Robinson dos Santos and Elke Elisabeth Schmidt (eds.), *Realism and Antirealism in Kant’s Moral Philosophy: New Essays*, Berlin/Boston: De Gruyter, 197-220.
- Stern, Robert. 2010. “Moral Scepticism and Agency: Kant and Korsgaard.” *Ratio* 23(4): 453-474.
- 2012. *Understanding Moral Obligation: Kant, Hegel, Kierkegaard*, Cambridge: Cambridge University Press.
- Stratton-Lake, Philip. 2013. “Morality and Autonomy,” in Sensen 2013a: 246-61.
- Timmermann, Jens. 2007. *Kant’s Groundwork of the Metaphysics of Morals: A Commentary*, Cambridge, UK: Cambridge University Press.
- Wood, Allen W. 2008. *Kantian Ethics*, Cambridge, UK: Cambridge University Press.