

Introduction

Stefano Bacin and Oliver Sensen

Kant is widely conceived to be the “inventor” of moral autonomy (cf. Schneewind 1998:3). However, what is autonomy, and how and why did he develop this conception? The aim of the proposed volume is to answer these questions.

1. Kant’s conception of autonomy

As a first approximation, one can say that Kant introduced the notion of autonomy in order to solve the problem of how unconditional moral obligation is possible. In his earliest writing on moral philosophy in 1764, Kant notes “how little even the fundamental concept of *obligation* is yet known” (UD 2:298). How is it possible that one ought to do (or refrain from) something, even if one is not inclined to do so?¹ Kant’s mature answer, from the *Groundwork of the Metaphysics of Morals* (1785) onwards, is that this is possible if the moral law that obliges originates in the agent’s own reason: Autonomy is self-legislation (cf. G 4:431; M II 29:629). In its final form this means that (1) reason gives the content of the moral law; it says what is morally right; (2) reason makes the law obligatory; reason itself is the ground why the law is valid; and (3) reason by itself can motivate an agent to comply with the moral law.

It is a more controversial issue what the ‘self’ in ‘self-legislation’ refers to: Does Kant mean, for instance, that (i) an empirical self, i.e., an agent how she is conscious of herself in introspection, deliberately adopts a rule, e.g., a New Year’s Resolution? (ii) Does ‘self’ rather

¹ On Kant’s concept of obligation, see Sensen 2015a.

refer to a noumenal self, i.e., a part of oneself that is not accessible in introspection, but exists outside time? (iii) Alternatively, does it merely mean that reason gives the moral law *by itself*, or (iv) that it gives the law *to itself* (cf. Sensen 2013a: 1-3)? It is unlikely that ‘self’ refers to an empirical agent, since this would not provide unconditional moral obligation in Kant’s sense. If an agent willingly binds himself, he could always unbind himself, for “the one imposing obligation ... could always release the one put under obligation ..., so that ... he would not be bound at all to a duty he lays upon himself.” (MS 6:417). Instead, the claim seems to be that pure reason is *by itself* the source of the moral law, i.e., it does not need any outside help. This can be compatible with the view that reason also gives the law to itself, and that reason is not an empirical, but a noumenal capacity; however, the emphasis seems to be on the independence of reason’s lawgiving from empirical influences. In its core meaning, then, autonomy is the “*lawgiving of its own* on the part of pure and, as such, practical reason” (KpV 5:33).

Kant’s notion of autonomy is even more complicated than that because Kant seems to use ‘autonomy’ at three different levels, or in three different contexts (cf. Sensen 2015b: 204f):

(A) At one level, autonomy is a “property” or mode [*Beschaffenheit*] of the will (cf. G 4:447, 440). As such Kant argues that our rational capacities are the source of a priori laws:

In regard to the faculties of the soul in general, insofar as they are considered as higher faculties, i.e., as ones that contain an autonomy, the understanding is the one that contains the **constitutive** principles *a priori* for the **faculty of cognition** ...; for the **feeling of pleasure and displeasure** it is the power of judgment ...; for the **faculty of desire** it is reason, which is practical without the mediation of any sort of pleasure. (KU 5:196f)

The idea is that fundamental laws, in morality the moral law, lie a priori in our rational capacities: “Pure reason is practical of itself and gives (to the human being) a universal law

which we call the *moral law*.” (KpV 5:31) These laws are not something reason discovers as an independent reality, but “with complete spontaneity it makes its own order according to ideas” (KrV A548/B576), “as soon as we draw up maxims of the will for ourselves” (KpV 5:29). At one level, therefore, autonomy describes the source of the moral principle (cf. Sensen 2013c).

(B) At a second level, Kant argues that this account of autonomy is a requirement for any moral theory (cf. G 4:441-5). If a theory does not see the source of the moral law a priori in one’s own reason, the opposite, heteronomy, results, and the theory cannot yield unconditional obligation: “*heteronomy* of choice, on the other hand, not only does not ground any obligation at all but is instead opposed to the principle of obligation” (KpV 5:33). Alternative theories yield heteronomy, according to Kant, because they are dependent upon an empirical interest, such as a feeling: If the law does not arise from one’s own will, then “this law had to carry with it some interests by way of attraction or constraint, since it did not as a law arise from *his* will” (G 4:432f). But in that case I ought not to do something unconditionally, i.e., independently of what I want, but “I ought to do something *because I will something else*.” (G 4:441) However, if the reason one acts on a principle is an empirical interest such as hope or fear, then one lets oneself be governed by the laws of those natural feelings, and then “it would, strictly speaking, be nature that gives the law” (G 4:444).

The reason heteronomy cannot yield unconditional obligation is that feelings are relative and contingent, and therefore cannot ground universal and necessary laws: “From the feeling of a sensation that may be different in every creature, no generally valid law can be derived for all thinking beings, and this is how the moral principle must be constituted.” (M II 29:625; cf. G 4:442; KpV 5:26).

(C) Finally, Kant seems to name one formula of the Categorical Imperative the “principle of autonomy” (G 4:433). As such it runs: “act as if your maxim were to serve at the same time as a universal law (for all rational beings).” (G 4:438). The specific aspect of this formula is “to indicate in the imperative itself the renunciation of all interest” (G 4:431). This is done by considering one’s maxim not from one’s subjective, but an impartial standpoint, so “that just because of the idea of giving universal law *it is based on no interest*” (G 4:432).

2. The aim of this volume

The contributors to this volume discussed the above aspects of autonomy through various meetings. The group was initially invited by Andrews Reath and Jens Timmermann to read the primary sources together, to discuss the different layers of Kant’s notion of autonomy, and to give comments on the first drafts of the authors’ chapters. This does not mean that each author agrees with the above account of autonomy. Kant talks about different aspects in different writings, and an author might make the case that autonomy really should just be one aspect (or a few) of the ones mentioned above. However, the aim of this volume is not to reach a full agreement by all authors on every aspect, but to have an in-depth look at the different writings, and to determine which aspect of autonomy Kant puts forth in that work or in a specific context. However, together, these contributions sketch the windy road on which autonomy emerged in Kant’s moral philosophy.

The aim of this volume is also not to find one moment, influence, or (psychological) reason why Kant developed his notion of autonomy. While he may be the creator of moral autonomy, Kant himself did not invent the word ‘autonomy.’ Kant’s notion of self-legislation

draws upon a phrase in Paul's *Letter to the Romans* (2, 14f) that had often been commented upon by previous moral philosophers, including Wolff.² But it is our impression that Kant did not simply get the word from another author. It was only when he was 60 years old, writing his *Groundwork of the Metaphysics of Morals*, when he first introduced the term. He read Rousseau much earlier in his career, and the word 'autonomy' does also not appear in the textbooks by Achenwall or Baumgarten, with which Kant used in his lectures at the time, or in Garve, to which the *Groundwork* can be seen as a response.

It seems more likely that Kant was simply aware of the political use of the term, which was prominent at the time (cf. Feil 1987). Prior to Kant, autonomy was primarily a political concept that described the source of the laws of political communities. If a city-state, for instance, received its laws from a mother state, it was therefore a colony and did not govern itself. But if the city gave itself its own laws, it was labeled as autonomous. The Stoics had already used the political term in relation to an individual, as appears in Epictetus's characterization of freedom as "something independent and self-governing (*αὐτόνομος*)" (*Discourses*, IV.1, 56).³

Therefore, the aim of the volume is to trace the philosophical reasons that prompted Kant to put forth his conception of moral autonomy, and to trace these steps in the chronological order in which Kant wrote his works. This explains the order of the chapters.

3. Chapter summaries

² Cf. e.g. Wolff 1720: § 24. For other uses of the Pauline passage prior to Kant see Bacin 2013: 61.

³ On the Stoic conception see Cooper 2003.

In the first chapter, Heiner Klemme examines the initial problem that prompted Kant to develop his account of autonomy: How is moral obligation possible? Klemme's paper unfolds in three steps: He first characterizes the prominent debates on obligation around the time of the young Kant. In a first step, Klemme analyzes the rationalist views put forward by Wolff and Pufendorf, and brings out the problems that plague these views. In a second step, Klemme looks at the attempts of a young Wolffian, Johann August Eberhard, as well as Lord Kames who wanted to integrate the emphasis on feelings by recent British Moral Sense Theories. Finally, in a third step, Klemme indicates how Kant ultimately reacts to the two lines of approaches, and tries to mediate between them. Klemme argues that Kant – against Wolff – will shift from a material a posteriori principle to a formal a priori one, and that the mature Kant introduces respect for the moral law as an alternative to the feeling of Moral Sense Theories. Klemme's chapter, therefore, describes the initial problem, and the philosophical motives that prompted Kant to embark on the journey to solve the problem of moral obligation two decades later.

Susan Meld Shell, in Chapter 2, focuses on Kant's first writings in the 1750s and 1760s, in order to analyze which of the three main elements of autonomy are present early on: (i) that reason formulates the content of the moral law; (ii) that reason provides the binding force of the law; (iii) and that reason provides the motivation to comply with the moral law. Shell starts with Kant's earliest writings on cosmology, and demonstrates how Kant's engagement with Wolff, Crusius, and later Rousseau formed his early views on the subject. She argues that – while Kant already displays some element of his mature philosophy, such as the emphasis on categorical imperatives and a universalization procedure, the subordination of religion to morality as well as the independence of selfish impulses – he does not yet identify the source of the moral law's

content. This includes the obligatory and/or necessitating character of the law, or the incentive to obey it, in the lawgiving activity of pure practical reason. However, the elements that Kant did develop early on become essential features of his mature treatment of autonomy as “the fundamental principle of morality.”

The next two chapters try to shed light on the emergence of autonomy in Kant’s writings by looking at his critique of alternative moral views. In his mature philosophy, Kant will charge all alternative theories other than his own of yielding heteronomy and being unable to account for moral obligation. Kant’s criticisms, which change little between the 1760s and 1785, show his dissatisfaction with previous accounts of how moral obligation is possible, and shed light on what prompted him to develop the three features of autonomy.

In chapter 3, Stefano Bacin examines Kant’s critique of rationalist theories of morality, and focuses on Kant’s lectures on practical philosophy from the 1760s to the time of the *Groundwork*. Bacin first explains Kant’s taxonomy of moral theories, and portrays the perfectionism and theological conceptions of Wolff and Crusius respectively that Kant argues against. Bacin then, second, considers the arguments that Kant deploys against rationalism and their similarities to previous debates. Third, Bacin argues that Kant’s discussion of the rationalist views does not amount to an unqualified dismissal of it, but to a differentiated attempt at working out a novel rationalist position. It takes from the theological views the insight that moral obligation has to be generated by an act of a will, and from perfectionism the idea that the moral law is a non-positive, necessary law. Bacin’s paper explains how Kant’s requirements for what counts as a moral theory guided the development of Kant’s view about the proper source of the moral law and its binding force.

Georg Mohr, in Chapter 4, then analyzes the other part of Kant's critique of alternative theories: his objections against empiricist moral theories, and in particular any theory that grounds morality on feelings. The text bases for this chapter are the mainly unpublished notes and quotations from lectures dealing with moral empiricism and moral sense philosophy from the period between 1762 and 1785. Mohr observes that Kant, after having sympathized with moral sense philosophy, especially between 1762 and 1766/68, adopts a strategy that combines rationalism and empiricism during the 1770s. At that point, Kant argues that a rational moral principle has to be supplemented by a moral feeling in order to be able to motivate an agent. While Kant keeps this general structure during his mature philosophy, Mohr points out that in the end even the moral feeling seems to fall more on the rational side. Mohr's chapter therefore provides a detailed account of Kant's arguments against empiricist moral theories, and portrays the history and development of one emerging element of autonomy: that reason can motivate to abide by the moral law.

In chapter 5, Oliver Sensen focuses on Kant's *Lectures on Ethics* of the 1770s to determine which of the three central elements of autonomy Kant held during his silent period: (i) reason provides the content of the moral law; (ii) reason makes moral laws obligatory, and (iii) pure reason can by itself be practical. Sensen's conclusion is that Kant essentially held his mature position on the first element by 1770, and indicates the second by the middle of the 1770s – even if he does not yet state either of them unequivocally. However, Kant does not yet conceive of his mature views on moral motivation. While he holds that reason *does* have a moving force of its own, he also argues that by themselves our rational capacities are not strong enough to overcome inclinations. Since Kant does seem to have two aspects of autonomy by the middle of the 1770s,

but not yet the third, this supports the claim that moral motivation is an important part of Kant's conception of autonomy.

In chapter 6, Jens Timmermann places emphasis on the question of moral motivation, and examines two sources that were written just before Kant embarked upon the project of the *Groundwork of the Metaphysics of Morals* in 1784, the first *Critique's* 'Canon of Pure Reason,' as well as the Pölitz lectures on natural theology. In the 'Canon,' Kant assumes that moral laws are unconditionally valid. Human beings acknowledge that their own well-being is not good if they are not morally good. By itself, however, this insight cannot motivate them to act morally unless they are assured that in performing moral action, they do not ultimately sacrifice their own welfare. The laws of morality, Kant holds at that time, would not be incentives if they did not carry with them promises and threats issued by God. Timmermann's chapter analyzes at length the final missing piece of autonomy: the idea that reason can by itself motivate an agent to act. In doing so, Timmermann also examines many of the features that make autonomy possible, and that are necessarily tied up with Kant's mature conception of autonomy.

Eric Watkins, in chapter 7, examines the next chronological step in Kant's writings, the *Prolegomena* of 1783, and considers whether Kant's theoretical philosophy could provide the missing link to Kant's mature conception of autonomy. Watkins examines, in a first section, how, according to the *Prolegomena*, reason, in the guise of the understanding, makes nature possible in the formal sense by prescribing (the lawfulness of) laws to nature. In a second section, Watkins analyzes points of similarity and contrast between the ways in which reason prescribes laws in theoretical and practical contexts. In particular, Watkins compares the definition of lawgiving in both areas, as well as the three main aspects of autonomy. What these

considerations reveal is the close link between Kant's theoretical and practical philosophy on the view that reason, broadly construed, dictates the content of laws, and Watkins also shows the similarities on Kant's mature views that reason makes laws obligatory, and motivates. Despite important points of contrast, the extensive parallels between Kant's theoretical philosophy of 1783 and his mature practical philosophy make a strong, even if not definitive case in favor of the *Prolegomena* forming a central part of an explanation of the emergence of autonomy in the *Groundwork*.

In chapter 8, Marcus Willaschek discusses the 'Introduction' to the lecture *Naturrecht Feyerabend*, which Kant held in 1784, a few months before he sent the *Groundwork* to the publisher. The lecture notes contain the first securely dated occurrence of the term 'autonomy' (and the first occurrence in the context of moral philosophy proper) in Kant's documented work. Willaschek first determines the aim of the Introduction, and analyzes Kant's argument that freedom must be a law to itself. Willaschek then carefully examines the passages in which Kant uses the term 'autonomy,' and compares them to the *Groundwork* conception of autonomy. Willaschek argues that in the 'Introduction' Kant does not talk about the source of the Categorical Imperative, but at the level of the restriction of freedom by particular laws, e.g., a law not to lie. Furthermore, while Kant argues that moral laws need to be universal, his emphasis is not that one legislates for all rational beings, but only that one has to restrict one's own freedom. On the other hand, Kant already displays his mature view on moral motivation, and puts forth respect for the law as a non-empirical incentive.

Since autonomy is originally a political notion, Pauline Kleingeld, in chapter 9, examines whether Kant's political writings prior to the *Groundwork* shed light on Kant's account of moral autonomy. In

particular, Kleingeld focuses on the main text of the *Feyerabend* lectures notes on natural right (1784). For Kant's account of the two-tiered structure of political legislation parallels his distinction between two levels of moral legislation: on one level, the constitution of the state and the Categorical Imperative, which both provide normative procedures that should be followed, and, on another level, particular positive state laws and maxims that need to be in accordance with the higher level. On Kant's view, both the political constitution and the Categorical Imperative are a priori principles of reason. Given that the supreme moral principle has its source a priori in reason, and given that Kant, in the *Groundwork*, identifies the will with practical reason, then any substantive moral duties have their normative ground in the will itself. The analogy with the origin of laws in the political realm therefore explains why Kant finds the political notion 'autonomy' apt in his moral philosophy as well.

In chapter 10, Andrews Reath presents Kant's mature conception of autonomy as it emerged in the *Groundwork* (1785). Reath first argues that Kant sees autonomy as the solution to the philosophical problem of obligation or the authority that common moral thought assigns to moral requirements, and identifies Kant's core idea that the nature of rational volition is the source of its fundamental principle. Reath then presents Kant's argument, according to which any other moral conception would yield heteronomy, and that it is a requirement of any theory of unconditional obligation that the rational will is the source of the moral principle. Finally, Reath links this conception to some of its preconditions and implications, for instance, that the will is practical reason, and that moral considerations should eclipse any other concerns. Reath also argues why respect for the moral law as the right moral motivation has to be part of Kant's conception of autonomy, for if the law should be unconditionally valid, the incentive cannot be contingent inclinations. While the earlier chapters focused on individual elements of Kant's conception of autonomy, Reath's chapter tries to show how these different elements form one unified system.

While the main emphasis of this volume is the emergence of Kant's notions of moral autonomy, Kant himself later goes on to ascribe autonomy to all three higher faculties of the soul: the faculty of understanding, the faculty of judgment, and the faculty of desire. In the final chapter, Pierre Keller looks at Kant's conception of autonomy as it progresses after the *Groundwork* in the three *Critiques* and the *Anthropology* lectures. Keller argues that in all three areas: cognition, aesthetic judgment, as well as morals, Kant uses the same account of autonomy: that one's own reason is the source of principles. Keller then argues that this does not lead to egoism, but that the lawgiving – in virtue of being a priori – always has a universal and pluralist aspect to it. Keller also argues that the emphasis on the a priori source in reason does not mean that autonomy is wholly detached from a human beings' sensibility. This is because in all three areas, cognition, the aesthetic feelings, and the moral feeling of respect, sensibility is shaped by an a priori principle.

Taken together, the contributions to this volume paint the following picture: Kant was prompted to develop his conception of autonomy by the question of how unconditional moral obligation is possible. In order to answer this question, and in light of the failures of alternative theories, Kant developed the view that reason is the source of its own moral principle. Since the law should be unconditionally valid, this leads to three further aspects: Reason gives the content of the law, it makes the law obligatory, and by itself it is able to motivate an agent. Together, the chapters of this volume trace the long, multifaceted, but progressive way of how Kant invented moral autonomy.