

# KANT ON MORAL AUTONOMY

EDITED BY  
OLIVER SENSEN

*Tulane University*



**CAMBRIDGE**  
UNIVERSITY PRESS

# Contents

<i>List of contributors</i>	page ix
<i>List of abbreviations</i>	xi
Introduction <i>Oliver Sensen</i>	I
PART I KANT'S CONCEPTION OF AUTONOMY	13
1 Kantian autonomy and contemporary ideas of autonomy <i>Thomas E. Hill, Jr.</i>	15
2 Kant's conception of autonomy of the will <i>Andrews Reath</i>	32
3 Vindicating autonomy <i>Karl Ameriks</i>	53
4 Progress toward autonomy <i>Paul Guyer</i>	71
PART II THE HISTORY AND INFLUENCE OF KANT'S CONCEPTION OF AUTONOMY	87
5 Transcending nature, unifying reason: on Kant's debt to Rousseau <i>Richard Velkley</i>	89
6 Kant and the "paradox" of autonomy <i>Susan Meld Shell</i>	107
7 Autonomy in Kant and German Idealism <i>Henry E. Allison</i>	129

8	Autonomy after Kant <i>J. B. Schneewind</i>	146
9	Personal autonomy and public authority <i>Katrin Flikschuh</i>	169
PART III THE RELEVANCE OF KANT'S CONCEPTION FOR CONTEMPORARY MORAL PHILOSOPHY		191
10	Moralized nature, naturalized autonomy <i>Heiner F. Klemme</i>	193
11	Autonomy and moral regard for ends <i>Jens Timmermann</i>	212
12	"A free will and a will under moral laws are the same" <i>Dieter Schönecker</i>	225
13	Morality and autonomy <i>Philip Stratton-Lake</i>	246
14	The moral importance of autonomy <i>Oliver Sensen</i>	262
	Postscript: heteronomy as the clue to Kantian autonomy <i>Onora O'Neill</i>	282
	<i>References</i>	289
	<i>Index</i>	298

# *Introduction*

*Oliver Sensen*

The concept of moral autonomy is one of Kant's central legacies for contemporary moral thought. Today autonomy – understood as individual independence from coercion in making decisions – is variously considered to be a capacity all normal adults have, a goal one should strive for, and especially a moral right one can claim from others (cf. Hill 1991: 44–51). To violate a person's autonomy is considered to be a serious moral offence. Autonomy is put forth as a fundamental principle of medical ethics (cf. Beauchamp and Childress 2008), and sometimes even as the justification for human rights (cf. Griffin 2008: 151–52.). For an explanation and justification of this idea, scholars commonly refer to Immanuel Kant, the “inventor” of moral autonomy (cf. Schneewind 1998: 3).

But how does Kant conceive of autonomy? What is the relation of Kant's conception to contemporary notions of autonomy? And what exactly is the significance of Kantian autonomy for morality? These three questions are at the heart of this volume, which brings together new contributions from many distinguished experts on Kantian autonomy. In their thoughtful engagement with these important questions, the fifteen essays in this volume shed light both on the history of an important contemporary idea, and on the philosophic grounds for taking autonomy seriously.

The essays in the first part of this volume focus on clarifying Kant's conception of autonomy. Given the complex and multi-faceted character of his conception, it is not surprising that the contributors' answers to these questions do not always agree. For while Kant presents moral autonomy not merely as individual independence but literally as self-legislation (*Selbstgesetzgebung*), it is far from simple to pin down precisely which ‘self’ and which ‘legislation’ Kant designates with this term, and what is the precise relation between these two components of autonomy.

For instance, in the context of autonomy, does Kant use ‘self’ to refer to an empirical self, the way we know ourselves in deliberating and making

decisions? Does he refer, in other words, to a conscious decision to adopt a particular moral rule? This is one way of interpreting ‘self’ in the literature, but the challenge for this interpretation is to explain how a law that is legislated by an empirical self could be binding, since “the one imposing obligation ... could always release the one put under obligation” (*TL* 6:417). Indeed this obvious problem tends to make the very idea of self-legislation sound paradoxical. Second, then, perhaps in speaking of self-legislation Kant refers to legislation not by an empirical self, the self one knows through introspection, but a transcendental or noumenal self; perhaps it is one’s own pure reason that legislates a law independently of one’s desires and inclinations. However, this interpretation faces the challenge that it is not clear how this self is to be understood metaphysically. Is this a plausible understanding of ‘self,’ or merely ‘panicky metaphysics’ (Strawson) to rescue moral obligation? As a third candidate, perhaps there is a way to understand ‘self’ that can avoid both of the challenges facing the previous two readings: The ‘self’ in self-legislation might refer not to a person – whether empirical or noumenal – but to ‘legislation.’ It might specify that the law to which Kant is referring is not conditioned upon something else, but is its own legislation. ‘Self’ would then qualify principles, not persons (cf. O’Neill 2003: 11–19).

In short, one can reformulate the question about Kant’s usage of ‘self’ in self-legislation by asking: the autonomy *of what?* Does Kant’s conception refer to the legislation of empirical persons, of pure reason, or of a specific kind of principles?

To shed light on this question, one might seek to clarify how Kant uses ‘legislation’ in connection with autonomy. But here too there is controversy in the literature. In other contexts Kant distinguishes two aspects of legislation: the formulation of the content of a law, and the authority or bindingness of the law. For instance, a parliament can formulate a law, but it is the head of state who signs it into law and thereby makes it authoritative (cf. *RL* 6:313). In which sense does Kant speak about self-*legislation*? The contributions to the first part of this volume explore different answers to this question.

The essays in the second part of this volume focus on the development of Kant’s conception of autonomy and its relation to contemporary conceptions. This issue is connected to the first question of how Kant conceives of autonomy; for if his understanding of self-legislation turns out to be very different from contemporary conceptions of autonomy, it will be difficult to draw a direct line between present-day appeals to the principle of moral ‘autonomy’ and the explanation and justification of this

principle offered by Kant. The contemporary usage of ‘autonomy,’ for instance, seems to focus on the decision of an empirical person. If it turns out that this is not Kant’s understanding, then how does his view relate to the contemporary understanding, and in which sense might Kant be an inspiration for present-day thought?

The third part of this volume focuses on the question of how relevant Kant’s conception of autonomy is to contemporary debates. If Kant’s conception turns out to be different from contemporary views of autonomy, why should it be relevant for thinking about moral issues today? Contemporary usage of autonomy is sometimes criticized for being overrated in its moral relevance. For instance, if one decides for oneself which course of action is right, this could yield morally good or neutral actions, but it might also lead to evil ones (cf. O’Neill 2003: 5–6). But if one argues that Kant’s notion of autonomy is different from contemporary conceptions, why is it morally significant? For instance, if Kantian autonomy is understood to refer to a principle that is not conditioned upon external elements, what is morally important about that? Even if Kant’s conception of autonomy is very different from contemporary conceptions, are there ways in which the study of Kant’s view of autonomy can inform and enrich contemporary debates?

These are the broad questions with which the fifteen essays in this volume engage. The remainder of this Introduction will offer a preview of the more specific concerns and claims of each contribution.

#### PART I: KANT’S CONCEPTION OF AUTONOMY

In [Chapter 1](#), “Kantian autonomy and contemporary ideas of autonomy,” Thomas Hill argues that contemporary notions of autonomy differ from Kant’s conception. Hill distinguishes three commonly recurring themes regarding autonomy in contemporary ethics: autonomy is often considered to be (1) a *right* to make one’s own decisions without undue interference by others; (2) a *capacity* to make decisions with due reflection and independence of mind; but also (3) an *ideal* of living an autonomous life. While these themes often appear in applied ethical discussions of informed consent in medicine, of paternalism in law, and of the aims of education, Hill notes that Kant addresses autonomy in the context of abstract and foundational questions of moral theory. For Kant, autonomy is a necessary presupposition of all morality, and the main features of rational beings with autonomy is that they (a) have a *will* as a form of causality of rational beings; (b) are *free in a negative sense*, as not being

determined by prior physical or psychological causes; and (c) are *free in a positive sense*, by being able to act on a law of pure reason, the moral law or Categorical Imperative. But while Kant's conception of autonomy differs from contemporary ones, Hill argues that it can – as the core of a framework of moral deliberation – inform more concrete answers in the three contemporary contexts of applied ethics. Regarding the first view of autonomy as a right (1), Hill notes that Kant's framework provides the idea of the Categorical Imperative, according to which deliberation about specific moral principles must find principles acceptable to everyone, and thereby affirm each person's right to govern his or her own life. Regarding the second view of autonomy as a capacity (2), Kant's framework affirms the existence of a *capacity* for self-determination and as something others should respect. Finally concerning the third view of autonomy as the *ideal* of an autonomous life (3), Kant's framework supports the idea that all human beings have a disposition to such a life and the presumptive reason to support it in a context-sensitive manner.

Chapter 2, "Kant's conception of autonomy of the will," focuses on the precise nature of Kant's claim that autonomy "is the property of the will by which it is a law to itself" (*GMS* 4:440). Andrews Reath gives a close analysis of what prompts Kant to introduce autonomy, of his notion of will, of the law involved, and of the sense in which will is a law to itself. Reath points out that autonomy is introduced by Kant specifically in order to explain the authority of moral principles in ordinary thought – that is, the fact that moral requirements are said to hold unconditionally. In discussing self-legislation, the law Kant talks about is then the moral law or Categorical Imperative. The categorical nature of moral laws cannot be achieved if they are based on an interest the agent has or on some feature of a potential object of volition in which we have an interest, for then the law would be conditioned. The unconditional authority of the moral law can be achieved only if the law arises from the will as pure practical reason. Reath carefully distinguishes between different aspects of the will in the wider sense – Kant's distinction between *Wille* and *Willkür* – and explains how the Categorical Imperative can arise from the will, or how the will is a law to itself. Reath explains that the will is a kind of causality particular to rational beings, in which the representations that guide a being's activity to realize its objects are based in reason. Reath describes the Categorical Imperative as the internal constitutive principle that arises from the nature of the will, and as an expression of what it is to exercise the will. The will has the formal aim of reasoning correctly from principles to practical judgment that carries a claim to universal validity, which

means that it is guided by the moral law. Based on this normative conception of autonomy, Kant can then speak about autonomy as a capacity of free will or a moral status in a derivative sense.

In [Chapter 3](#), “Vindicating autonomy: Kant, Sartre, and O’Neill,” Karl Ameriks sheds further light on how one should understand Kant’s conception of autonomy. He focuses on the two components of autonomy: *auto* as independence or negative freedom, and *nomos* as lawfulness or positive freedom. The first component could be described as ‘deciding on one’s own,’ and the second as ‘following a law of one’s own making.’ Ameriks distinguishes between an empirical and transcendental reading of each component, and warns against two misunderstandings of Kant. The first misunderstanding would be that ‘deciding on one’s own’ amounts to a radical existentialism; the second would be that ‘a law of one’s own making’ requires a panicky metaphysics, in alluding to a metaphysical self that demands uniform behavior. Ameriks refers to Sartre in discussing the first component of autonomy and for the second component he refers to O’Neill – in addition to many writings of Kant’s – to show that there are significant positions in-between a radical choice and authoritarian metaphysical self; and he argues that Kant occupies the middle ground between the two. On Ameriks’ account, for the *auto* component Kant does not refer to an empirical freedom from particular empirical forces, but to transcendental freedom as a will that is a unique cause wholly independent of empirical determination. Similarly, Ameriks explains that the ‘law of one’s own making’ component does not commit Kant to either an empirical choosing of rules, or to an authoritarian metaphysical self. Rather Kant emphasizes the absolute necessity of moral lawfulness, which cannot be founded on spatiotemporal grounds. ‘Of one’s own making’ is then a shorthand for emphasizing the structure of what is necessarily reasonable as such, not a description of an empirical process.

[Chapter 4](#), “Progress towards autonomy,” shows that there is also an important empirical side to Kant’s notion of autonomy. Paul Guyer argues that it does not reflect Kant’s full view to say that one either has or does not have autonomy – in the sense that one’s fundamental maxim is either (1) to be governed by the moral law only if it fits with one’s self-love, or (2) only to follow self-love if it complies with the moral law. Rather, Guyer notes that there is an empirical side to autonomy, in that compliance with the moral law can only be hard-won and achieved gradually. Guyer’s account makes use of the distinction between a negative and a positive aspect of autonomy – that is, of not being determined by outside forces, but of being determined by the moral law of one’s own reason. He



first argues that these are not two separate forms of autonomy, but rather two aspects of one unified form. An agent can achieve independence from outside determination only through positive autonomy; and since independence from outside determination can only be realized progressively, the same is true of the whole of autonomy. To further support his argument, Guyer discusses Kant's claim in the *Religion* that adopting a fundamental maxim concerning the subordination of morality to self-love, or the reverse, is an all-or-nothing decision; and he shows that it would at most capture an aspect of autonomy that is not given in experience. But on a psychological level, Guyer argues, being governed by the moral law is a gradual affair. This comes out in Kant's discussion of self-mastery in the *Lectures on Ethics*, as well as in his discussion of the aesthetic preconditions of morality in the *Doctrine of Virtue*.

PART II: THE HISTORY AND INFLUENCE OF KANT'S  
CONCEPTION OF AUTONOMY

In [Chapter 5](#), “Transcending nature, unifying reason: on Kant's debt to Rousseau,” Richard Velkley argues that Rousseau's influence on Kant's notion of autonomy goes deeper than the widely recognized kinship of autonomy with Rousseau's conception of the general will. Velkley notes that when Kant read Rousseau at about age forty, it led to a fundamental reorientation in Kant's conception of philosophy, a reorientation involving his views on nature, reason, desire, freedom, and history. Rousseau's works made Kant aware of a crisis in the intellectual life of Europe. On Rousseau's account, human life is burdened by luxury, vanity, and factitious desires, and Kant accordingly searches for a firm standpoint to counter these artificial desires. He finds it in freedom, whereby the factitious desires can be corrected by an opposing principle that rests in reason beyond nature. Human reason has an immediate awareness of the injustice of servitude, which includes foreign influences as well as luxurious desires. Therefore Velkley notes that most of the elements of Kant's mature notion of autonomy can be found in his thinking as early as his *Remarks on the Observations on the Feeling of the Beautiful and Sublime* (1764–65), even if he has not yet worked out the internal connection of the different elements and has not developed a motive for adopting the standpoint of reason. Velkley then argues that these early thoughts were not confined to Kant's moral philosophy, but are also of central importance for the theoretical use of reason; and Velkley traces the development of these thoughts from the early *Remarks* to Kant's mature writings.

Chapter 6, “Kant and the ‘paradox’ of autonomy,” traces the development of Kant’s conception of autonomy from his earliest writings on cosmology to the *Groundwork*. Susan Shell describes this development as beginning from Kant’s early question of how separate entities as substances can stand in external relations, and from the paradoxes to which this gave rise. She shows that Kant grapples with this and related moral questions throughout his theoretical and practical writings, as well as his lectures, until he finds a solution in the *Groundwork*. Shell argues that Kant’s conception of autonomy, along with the related concept of a kingdom of ends, allows him in the *Groundwork* to conceive of the possibility of a community of substances that are related not only externally, but also through a reciprocal lawgiving. It is then possible to explain how external relations among substances-in-themselves are possible without positing a ground that is not accessible to human reason. And whereas the paradox of how material substances can interact is merely a metaphysical question, what matters is that one think pluralistically, as one citizen of the world.

In Chapter 7, “Autonomy in Kant and German Idealism,” Henry Allison presents Kant’s conception of autonomy and the modification and criticisms of it by Fichte, Schiller, and Hegel. Allison first describes in a concise way Kant’s conception of autonomy, its importance for morality, and Kant’s attempts to show that human beings have autonomy of the will. Allison then discusses the attempts of Fichte and Schiller to clarify and repair Kant’s conception, and Hegel’s attempt at a radical transformation of it. Fichte, Allison notes, tries to expand Kant’s notion of autonomy into a conception of the self or I. For Fichte, the autonomy of the self is the ultimate starting point of all philosophy, theoretical and practical; everything is to be explained in terms of the self. Autonomy in this sense is not a property of the will, but a pre-conscious activity of self-determining. As Allison observes, Fichte thereby turned Kant’s moral philosophy into an even more rigoristic direction by not allowing any room for inclinations. In contrast, Allison notes that Schiller tries to temper the rhetoric of Kant’s rigorism by emphasizing inclinations. While the imperatival form of Kant’s moral law makes it appear to be a foreign law, Schiller argues that true freedom is realized when there is a harmony between rational and emotional factors, or an inclination to duty. Allison notes that Hegel goes beyond Schiller in arguing for the need of a full integration of rational law and inclinations. According to Hegel, Kant’s notion of autonomy mentions merely one aspect of the will, which conflicts with the possibility of agency; in Hegel’s view, action presupposes an awareness of an ‘other,’ and this ‘other’ stems from particular interests.

In merely emphasizing the abstract law, Hegel claims, Kant leaves out an essential moment; only through a synthesis of both elements can rational self-determination be attained. For Hegel, then, autonomy becomes an ideal state to be attained.

Chapter 8, “Autonomy after Kant,” sheds light on the reception of Kant’s views of autonomy in the English-speaking world. J. B. Schneewind surveys decades of literature on autonomy, and he brings to light the reasons for the resurgence of autonomy over the last 40 years, after it had largely receded from view during the roughly 100 years following the works of John Stuart Mill. In particular, Schneewind discusses five areas which were conducive to the renewed interest in autonomy: (1) new ideas on free will and the philosophy of action (e.g., Frankfurt and Dworkin); (2) medical ethics and bio-ethics (led by Beauchamp); (3) feminist debates about sexism and women’s liberation (e.g., by Gilligan and Held); (4) liberalism and its critics in political thought (e.g., Rawls); (5) Kantian commentary and the revival of Kantian ethics. Schneewind discusses the relationship of the notions of autonomy employed in these different areas to Kant’s conception of autonomy, and he defends Kant’s conception against two recent challenges: an attack on the possibility of an autonomous self, and the claim that such a self is historically contingent and a social construction.

In Chapter 9, “Personal autonomy and public authority,” Katrin Flikschuh examines the influence of Kant’s conception of autonomy on political thought. Personal autonomy – as an individual’s capacity to lead his or her own life – has become the primary value of contemporary liberalism, and leads to the view that it is desirable for persons to collectively arrange their common affairs as they see fit. Kant is often seen as the inspiration for these views, but against this, Flikschuh argues on systematic grounds that personal autonomy does not by itself explain why one should take an interest in other people’s autonomy; and on textual grounds, she argues that Kant’s view of public legislation differs from that of contemporary liberalism. According to Flikschuh, Kant neither has a view of collective self-legislation – according to which each individual has a veto power over proposed legislation – nor does he have a co-legislative scheme – in which each individual has a voice, even if the outcome can overrule the positions of individuals, who lack veto powers. On Kant’s account of public legislation, rather, the sovereign’s deliberation is omnilateral, considering what is right for everyone, rather than being a private will that deliberates only for its own interests. Flikschuh notes that it is only in this way that a public will can have authority over private wills, in

Kant's view. The relation of legislator and individual citizens is then vertical and coercive, for Kant, while giving the individual freedom of the pen to point out where the legislator might have been mistaken.

PART III: THE RELEVANCE OF KANT'S CONCEPTION FOR  
CONTEMPORARY MORAL PHILOSOPHY

In [Chapter 10](#) Heiner Klemme brings out the importance of Kant's conception of autonomy for the understanding of nature as well as morality. In an essay entitled "Moralized nature, naturalized autonomy: Kant's way of bridging the gap in the third *Critique* (and in the *Groundwork*)," Klemme first links autonomy and its opposite, heteronomy, to two types of causality, namely freedom as *causa finalis*, and natural causality as *causa efficiens*. He then shows the importance of autonomy for understanding nature as Kant presents it in the *Critique of the Power of Judgment*. There Kant introduces the perspective of autonomy to explain nature as operating on purposes (*causa finalis*), in addition to a mechanical explanation of nature (*causa efficiens*). Although the autonomy perspective is merely a regulative one – reason judges *as if* nature has a purpose – this perspective is nonetheless necessary for understanding nature itself, as Klemme explains. It is needed when the mechanical explanation is not sufficient for the understanding of phenomena, as is the case with organized products of nature like living organisms. Without the autonomy perspective, reason could not detect necessity in all of nature: the difference between a living organism and a dead one would be a mere whim of nature to us. But Klemme notes that the autonomy perspective is also central to morality, on this view. If one looks at nature as if it has a purpose, the question arises whether it also has a highest purpose that is not a means to something else. This highest purpose seems to be human beings themselves, since they are the only beings on earth who can conceive of a system of ends. On Kant's account, therefore, in studying nature reason discovers itself as a faculty that has to be appreciated for its own sake. The autonomy perspective Kant presents is important for a description of nature and moral philosophy.

In [Chapter 11](#), "Autonomy and moral regard for ends," Jens Timmermann explores this moral importance of autonomy for Kant in grounding our status as ends in ourselves. He argues that, on Kant's account, autonomy – as the capacity to act on self-imposed laws – is a pure will, insofar as it does not need an external object for its determination, but is determined by the formal moral law. Timmermann argues

that it is this self-referential structure, the will that wills itself, that makes a human being an end in itself, according to Kant. Everyone possesses this pure will no matter how morally bad he is, and as a capacity it cannot be corrupted, even if one does not act on it. In addition, Timmermann presents the consciousness of one's own pure will as leading to a sense of awe and elevation that can motivate one to act morally; one has reverence for one's own status as an end in itself. Since one's own rational self does not want to be undermined by sensibility or foreign determination, one should – on the presumption of equality – grant that same status to others. Thus Timmermann argues that, on Kant's account, autonomy – one's capacity to act on the self-imposed moral law in contrast to inclinations – has supreme moral importance, in being the ground of what should be respected in others.

In [Chapter 12](#), Dieter Schönecker offers a close textual reading of the *Groundwork* passages in which Kant argues for the tight connection between freedom and morality. In an essay entitled “‘A free will and a will under moral laws are the same’: Kant's concept of autonomy and his thesis of analyticity in *Groundwork* III,” Schönecker explains that, as Kant sees it, to have free will is at the same time to be under the moral law. Schönecker first distinguishes two readings of this claim. Kant could hold (a) that in virtue of freedom, a perfectly rational being that is not under the influence of inclinations would be guided by the moral law as a descriptive law; or (b) that in virtue of freedom, a human being who possesses reason and inclinations is bound by the Categorical Imperative. Schönecker argues that Kant holds the weaker claim (a), and that it requires a separate argument to show that morality is binding for human beings. He then presents Kant's argument for (a) that relies on autonomy: freedom is a form of causality, and as a form of causality, it involves laws. These laws cannot come from outside, as Kant insists that this would be heteronomy. Therefore, Kant concludes, freedom of the will must be the will's quality of being a law to itself, i.e., autonomy. For Kant, autonomy is important in bringing out the link between freedom and morality.

In [Chapter 13](#), “Morality and autonomy,” Philip Stratton-Lake questions the tight connection that Kant draws between morality and autonomy, from the perspective of contemporary ethics and a reasons-for-action account. Stratton-Lake first presents two arguments against the importance of Kantian autonomy – understood as legislating for everyone. In legislating for everyone, Stratton-Lake argues, one would undermine *their* autonomy. The second argument considers possible defenses the Kantian could make. For instance, the Kantian could reply that one

should legislate for everyone in a way that they could consent to. But Stratton-Lake notes that there can be cases where a person *could* in principle consent to an action; however, if in fact he does not consent, one would violate rather than respect his autonomy. After considering several possible replies, Stratton-Lake offers a reasons-for-action account as an alternative to the Kantian emphasis on autonomy. On Stratton-Lake's account, morality is based on facts in nature, e.g., the fact that something causes pain, that explain why one ought to do something and give a reason to comply. The advantage of this approach is said to be that one can always give a reason for why one should perform a moral action, other than that it is commanded by the moral law, as Kant's account seems to imply. Stratton-Lake argues that the reasons-for-action account can preserve some of the features of morality that are important to Kant, such as the view that moral demands apply independently of what one wants, and that one would not be forced by the will of another person. But Stratton-Lake's account disconnects the necessary link Kant sees between autonomy and morality. If one follows a mistaken judgment about what is right, on Stratton-Lake's view, one acts autonomously, but not in the morally right way.

In Chapter 14, "The moral importance of autonomy," I attempt to spell out wherein Kant sees the moral significance of autonomy, as well as his argument why, for instance, a reasons-for-action account could not yield moral obligation on Kant's account. I first argue that although Kant regards empirical self-determination as important for one's happiness, mere choosing is not the reason why Kant considers autonomy of the highest moral importance: a self-determined scoundrel is even more abominable in our eyes than someone who performs an immoral deed unintentionally. Rather, I note that Kant understands autonomy as reason not only giving the content of the moral law, but also binding one to it; and I argue that Kant sees the significance of autonomy in the conclusion that it alone can generate moral obligation. He puts forth the bold claim that any other moral theory could only yield conditioned or hypothetical laws, but not an unconditioned or categorical one. Accordingly, I try to spell out Kant's argument for why moral theories that are based, for instance, on (1) desires, (2) positive laws, or (3) real normative properties could not ground unconditional obligation. Kant rules out desire-based accounts because desires and feelings are relative and contingent, and therefore cannot ground a universal and necessary moral law. Kant also groups moral theories based on positive laws (2) and real normative properties (3) together with desire-based accounts, since one would need a

desire to comply with a positive law, or a feeling to discern a real normative property. My aim in the chapter is to spell out Kant's premises for further consideration. One of these premises is that morality is and needs to be necessary and universal, and the chapter ends with a reflection on what Kant's arguments are for this claim.

In one way or another, the contributors to this volume have benefited from the work of Onora O'Neill on Kantian autonomy. In her postscript, "Heteronomy as the clue to Kantian autonomy," O'Neill argues that contemporary conceptions of autonomy are fundamentally different from Kant's notion. She locates the difference not in the claim that Kantian autonomy involves choosing freely and based on principles – for she notes that heteronomous actions can fulfill both of these requirements. Rather, O'Neill suggests that the difference lies in how these principles are justified, and accordingly in a difference in the scope of people for whom the relevant principle would be valid. As a clue to Kant's notion of autonomy, she examines what Kant describes as heteronomous principles. Heteronomous principles prescribe action merely conditionally; they are based on extraneous and ungrounded premises, for instance, desires, conventions, dogmatic worldviews, or specific accounts of happiness. Autonomous action, then, must merely meet a negative condition: that one's principles are not based on relative assumptions, but are fit for universal adoption. O'Neill notes that this explains the close link between autonomy and Kant's Categorical Imperative. Kant's view is thereby seen to be immune from a dilemma that can be raised against the concept of self-legislation. Kant neither emphasizes the decision of an empirical *self* – in which case one could not explain how one could be bound against one's will – nor does he smuggle in a *positive* conception of legislation – in which case there would be an *external* standard of morality. Thus in contrast to the contemporary emphasis on self-expression and self-interest, Kantian autonomy points to the ethical importance of action for which conclusive reasons can be given.

## CHAPTER I

# *Kantian autonomy and contemporary ideas of autonomy*

*Thomas E. Hill, Jr.*

### I INTRODUCTION

Kant argued that autonomy of the will is a necessary presupposition of all morality. His idea of autonomy is abstract, foundational, normative, and a key to his defense of the rationality of moral commitment. In contemporary ethics philosophers often appeal to ideas of autonomy more specifically, for example, as a moral right to make one's own decisions or as a psychological ideal of independent thinking and rational self-control. Kantian and contemporary ideas of autonomy are not the same, but how are they related? Arguably, Kantian autonomy, when applied realistically to common human circumstances, tends to support but also limit particular claims often made in the name of autonomy. Kantian autonomy is meant to be a crucial part of the moral point of view from which specific principles can be assessed, not an ideal of living independently of others. Derivative principles must take into account ways in which various physical and social conditions inhibit the capacity for self-determination and various values limit its exercise.

My discussion of these themes will be divided as follows. *First*, I summarize what I take to be main features of Kantian autonomy. This is an interpretation of what Kant says, text-respectful even if not always text-literal. Scholars and critics, I expect, will generally agree with these main points, though it remains controversial what more Kant intended. *Second*, I distinguish this Kantian conception of autonomy from philosophical ideas associated with existentialism, prescriptivism, and some feminist critiques. *Third*, I sketch three general conceptions of autonomy that appear in discussions of practical issues ('applied ethics'), noting that each is distinct from the Kantian conception and needs to be further defined and qualified. *Finally*, I suggest that Kantian autonomy, when considered as the core of a framework for moral deliberation, might serve a basis for developing and supporting more definite principles that correspond to



the three general conceptions of autonomy found in applied ethics. As a general way of orienting practical thought, the framework appropriately respects the need for context-sensitive facts and good judgment in discussions of specific practical issues, for example, informed consent in medicine, paternalism in law, and the aims of education.

## 2 AUTONOMY OF THE WILL IN KANT'S MORAL THEORY

Kant introduces the idea of autonomy as a part of his philosophical investigation into what we can know and reasonably believe a priori. He thought that analytical propositions, such as 'triangles have three sides' and 'lethal poisons can kill,' pose no problem because it would be self-contradictory to deny them. His critical philosophy focused instead on the question 'how is *synthetic* a priori knowledge possible?' For example, in *the Critique of Pure Reason* he inquired how can we know non-analytic propositions such as ' $5 + 7 = 12$ ' and 'Every event has a cause.' His investigation concluded that theoretical reason can validate such a priori synthetic propositions of mathematics and critical philosophy for use in empirical science, but theoretical reason can only establish the *possibility* of the soul, God, freedom, immortality etc. Synthetic a priori propositions about such matters cannot be established or even comprehended by science or traditional metaphysics that employs reason theoretically to understand existing natural or super-natural objects and events. Moral philosophy, however, is not about the world as we find it or imagine it to be, but about how we *ought* to act and what we *ought* to strive for. Questions in this domain require the practical use of reason. In his *Groundwork for the Metaphysics of Morals* Kant investigated the presuppositions of common moral consciousness and argued that the beliefs that we ought to maintain a good will and are subject to moral duties presuppose a fundamental synthetic a priori proposition, which I take to be this: "Any fully rational agent, when sufficiently guided by reason, would accept and follow the moral law, which is formally expressible as '*Act only on that maxim by which you can at the same time will that it should become a universal law*'" (*GMS* 4:419–21).<sup>1</sup> Kant then argued that this supreme moral principle, which can be represented in several other formulas, is the governing principle of any rational person with autonomy of the will (cf. *GMS* 4:440–45).

<sup>1</sup> Translations of the *Groundwork* are taken from Immanuel Kant, *Groundwork for the Metaphysics of Morals*, ed. Arnulf Zweig and Thomas E. Hill, Jr. (Oxford University Press, 2002).

The point to highlight here is that the context for Kant's use of the idea of autonomy of the will was a highly abstract discussion of fundamental questions of moral theory, not a scientific or metaphysical inquiry or a discussion of specific questions about moral and political rights, paternalism, the aims of education, or ideals of enlightened thinking. This is not to say that his idea of autonomy is irrelevant to these specific practical questions, but only that we must be cautious about moving from one to the other.

Kant's idea of autonomy plays a key role at several stages of his argument in the *Groundwork* (as well as later in the *Critique of Practical Reason*). The primary project of the first two sections of the *Groundwork* is to answer these questions: assuming, as we do, that we have genuine moral obligations and duties, how can we express the fundamental principle(s) from which they derive and how must we conceive of ourselves as the agents who are bound by them? Kant's answers, in brief, are that the fundamental moral principle is expressible in his several formulations of the moral law, including a principle of autonomy (cf. *GMS* 4:432–33), and we must conceive of ourselves, from the standpoint of practice, as rational agents with autonomy of the will. The common error of all previous moral theories, he argues, was the failure to understand and acknowledge that we are bound by moral obligations and duty only insofar as we are rational agents with autonomy as a property of the will (cf. *GMS* 4:440–44). These other theories present the fundamental requirements of morality in such a way that they would be binding only on the basis of hypothetical imperatives and interests that are not essential to moral agency. Their standards would be at best principles of heteronomy.

A further stage of Kant's argument in the *Groundwork* is developed in the third section, where Kant argues that our initial assumption that we are under moral obligation and duty is not an illusion, or so we can and must reasonably believe, because it is not just morality but all our practical thought that presupposes that we have rational autonomy. Autonomy of the will is inseparable from the negatively conceived freedom that is necessarily an aspect of our self-conception when we address the practical question, 'what ought I to do?' (cf. *GMS* 4:446–47). A rational person necessarily acts under the idea of this double-sided freedom. Even our theoretical inquiries, Kant suggests, require this self-conception (cf. *GMS* 4:450–53). Later, in the *Critique of Practical Reason*, the idea of autonomy continues to play a key role. Kant argues again that any rational will with autonomy has the moral law as its fundamental overriding standard and that this positive freedom is inseparable from freedom

negatively conceived, but now the necessity of conceiving of ourselves as free is inferred from our consciousness of moral obligation rather than the reverse (cf. *KpV* 5:29–33).

This is not the place to examine these basic claims and arguments in detail, but they are worth keeping in mind as the context for Kant's presentation and use of his idea of autonomy. The question to consider now is what are the main features that we attribute to moral agents when we conceive of them as rational persons with autonomy of the will? Here is a brief summary of what I take to be the main features.

*First*, we conceive of rational agents with autonomy as having *wills*. The *will*, according to Kant, is "a kind of causality that living beings have in so far as they are rational" (*GMS* 4:446). It is "a power to act in accord with the idea of laws – that is, in accordance with principles" (*GMS* 4:412). To say that human beings have a will, then, implies that they can make things happen intentionally and for reasons, where this is ultimately understood by reference to their policies, ends, and underlying rational principles. When I raise my hand on purpose, as we say, the relevant practical explanation is not just about the physical forces that caused my hand to rise but about my rationale for doing it. Presumably I have a conception of what I am doing and why, and this could be expressed as a *maxim* on which I am acting. The will is not, as it was for Hobbes, the last and strongest appetite in a deliberation that causes my body to move, nor is it, as it was for Hume, a peculiar sort of mental event ('impression') that occurs just prior to my acting.<sup>2</sup> Rather it is a power to cause changes in the world on the basis of a rationale, which if spelled out would include our beliefs, aims, policies, and an implicit idea of a relevant rational principle. In his later work Kant distinguishes two senses of 'will': a *legislative practical reason* (*Wille*) that prescribes universal rational norms of conduct (laws) and a *power of choice* (*Willkür*) by which we adopt our maxims and choose our particular actions. As imperfectly rational agents we do not always exercise our power of choice fully in accord with rational norms.

*Second*, the will of a person with autonomy of the will is *free in a negative sense*. That is, it is a kind of causality that "can be active, independently of alien causes determining it" (*GMS* 4:445). In other words, we conceive of a person with a negatively free will as able to act and cause events without the person's choices being causally determined by prior physical and psychological forces. Given that we make sense of a person's

<sup>2</sup> Thomas Hobbes, *Leviathan*, Part I, ch. vi: 33–34, and David Hume, *A Treatise of Human Nature* Book II, Part III, Section I.

will to do something by reference to the person's maxim, we should add that a person with a negatively free will has the capacity to act on maxims that are not based on that person's inclinations.<sup>3</sup> The rationale on which a person with a negatively free will chooses to act can be to pursue an inclination-based end from self-love, but it need not be so. Kant implies that negative freedom includes the capacity to act without always being motivated by instrumental reasoning in the pursuit of happiness. Our rationales for acting are not completely limited by the desires and interests that we find, as it were, given to us by nature.

*Third*, autonomy of the will is *freedom positively conceived (autonomy)*. Kant argues that even a negatively free will cannot be 'lawless,' and yet its negative freedom implies that it is not determined by empirical causal laws. Our inclinations and contingent interests, even in conjunction with hypothetical imperatives, do not provide us with 'laws' which are universal and necessary rational norms. So Kant argues that a negatively free will must be capable of making its choices by laws of another kind – that is, laws of pure practical reason (*GMS* 4:446–47). Autonomy is "the property the will has of being a law to itself (independently of any property of the objects of volition)" (*GMS* 4:440). Even imperfectly rational human beings recognize the laws of pure practical reason as their own standards, not imposed by any external authority or force and not based on inclination, moral sensibility, rational intuition, or divine command (*GMS* 4:441–44; *KpV* 5:35–40). By analogy with citizens in Rousseau's ideal state, we have a positive freedom (or 'moral liberty') not by being free from all laws but by being subject only to the authority of laws that we will for ourselves and others as members of a moral community in which every person is both author of the laws and subject to them.<sup>4</sup>

It should be emphasized that autonomy of the will is not simply the property of "being a law unto oneself" in an unqualified sense but rather the property of making and being subject to laws "*independently of every*

<sup>3</sup> The addition depends on a distinction between taking oneself to be able to act independently of empirical causes and taking oneself to be capable of acting independently of empirical motives. The distinction is relevant because someone might hold that particular choices are not fully determined by empirical causes but nevertheless we can only choose among options in which we have empirical interests. Kantian freedom implies that we are neither fully determined by empirical causes nor restricted to empirical interests. For further comments see my essay "The Kantian Conception of Autonomy," in Hill 1992: 85–86; originally published in Christman 1989: 91–105.

<sup>4</sup> The views are only analogous, however, because Rousseau was focused on political theory concerned with *coercive laws* of the state within a *limited jurisdiction*, and Kant places autonomy in his comprehensive moral philosophy concerned also with non-coercive ethical "laws" that guide individual choices within a moral community that includes all persons. See Jean-Jacques Rousseau, *Of the Social Contract* (1762 [1964]: esp. 1.viii).

*property belonging to the objects of inclination*” (GMS 4:440). Inclinations may vary from person to person and time to time, but the fundamental laws of a will with autonomy must be the same for all, as if legislated in a kingdom of ends that we can think of only by “abstracting from personal differences” (GMS 4:433–34). When at this most abstract level we ask what moral (mid-level) principles we would ‘legislate’ as rational persons with autonomy of the will, we must set aside personal preferences that distinguish one person from another. Even if certain inclinations are shared by all human beings, we cannot infer from that alone that it is rationally necessary to pursue or accommodate the things that would satisfy them. What fundamental choice-guiding principles rational agents with autonomy can and would ‘legislate’ as imperative for human beings must depend on their essential interests as rational autonomous agents who are subject to the human condition. Arguably, essential interests would include exercising, preserving, and improving one’s rational capacities under principles justifiable to all, but satisfying the desire for sweet-tasting food would be at most a derivative concern.

To have a *rational will with autonomy* (as described here) is, primarily, to have certain general capacities and dispositions to act and cause events in accord with our ideas of rational principles without being determined by non-rational (‘alien’) factors. Sometimes, however, we speak of a person’s *particular will to do a specific kind of act on a certain occasion*. These are ‘choices’ based on maxims (‘subjective principles’) that may or may not be morally justified. When one bases one’s particular choices on the moral law, one expresses one’s nature as a rational person with autonomy of the will. When one acts immorally, one’s particular choices conflict with the principles that one cannot help but see as one’s own rational standards.<sup>5</sup> Even these bad choices are the choices of a person with freedom of the will – negative freedom and the capacities and dispositions of rational autonomy.<sup>6</sup> Immoral choices do not *express one’s autonomy* but

<sup>5</sup> Some commentators suggest that, if we are rational agents with autonomy of the will, every act of will (such as the choice of a specific maxim) must be implicitly guided by or based on the principle of autonomy, that is, the moral law. See, for example, Engstrom 2009: chs. 4–6, 98–183, and earlier Wolff 1973. The interpretation presented here, by contrast, takes autonomy to imply our capacity and disposition to acknowledge and ‘impose on ourselves’ the moral law *as an overriding rational standard* but not that every choice is an implicit attempt to follow the moral law. When we act immorally, we have a conflict in our will, but we are still persons with autonomy of the will whose choices are, in a sense, exercises of our free power of choice (*Willkür*).

<sup>6</sup> By “rational autonomy” I mean *rationality* and *autonomy of the will* in the Kantian sense explained above. As Onora O’Neill notes, the term “rational autonomy” in contemporary discussions can mean something more limited. See her helpful contrast between Kantian and contemporary conceptions in her 2003.

this does not mean (as some say) that the choices are unfree or ‘heteronomous’ in a sense implying that one acted blindly, for no reason, as a mere product of empirical forces, without the capacity to know and follow the principles of rational autonomy.

3 KANTIAN AUTONOMY AS DISTINCT FROM OTHER  
INFLUENTIAL CONCEPTIONS

In previous essays I have distinguished Kant’s idea of autonomy from ideas of freedom and autonomy that are central in the work of other philosophers.<sup>7</sup> One might suppose, for example, that for Kant autonomy was just what earlier Rousseau had called “moral liberty” in contrast with “natural liberty” and “civil liberty.” For Rousseau, moral liberty – “obedience to the law one prescribes to oneself” – becomes possible only within an ideal state in which the laws are backed by the general will of the people.<sup>8</sup> Here the citizens have a basic commitment, expressed in the social contract, to subordinate their private wills to the general will, which is always unflinchingly for the common good. Even in the ideal state individuals are sometimes tempted to pursue their private interests contrary to the laws of the general will, and then the state may compel them to obey. In this case, however, they are merely being “forced to be free” because they are only compelled to obey laws that they, as citizens sharing in the general will, prescribe to themselves.<sup>9</sup> This conception of ‘positive liberty’ has been severely criticized as an unworkable or dangerous model for political theory,<sup>10</sup> but the relevant point for now is that, despite some affinities, it is very different from the autonomy of the will that Kant took to be fundamental for all aspects of moral philosophy (including virtue as well as justice). Rousseau aimed to explain what could justify the use of coercive state power, and he attributed “moral liberty” only to public-spirited citizens under ideal social conditions. Rousseau’s citizens ‘prescribed laws to themselves’ by a public process of voting, not by Kantian ‘inner legislation,’ and the state laws were binding only within a limited jurisdiction, not (as Kant’s moral laws are meant to be) for all humankind.

<sup>7</sup> For example, “The Kantian Conception of Autonomy,” cited above, “Autonomy and Benevolent Lies” (Hill 1984; reprinted in Hill 1991: 19–51), and “The Importance of Autonomy,” in Kittay and Meyers 1987: 129–38 (reprinted in Hill 1991 and, in German translation, “Die Bedeutung der Autonomie,” in Ameriks and Sturma 2004: 178–89).

<sup>8</sup> Rousseau 1762 [1964]: i.viii, third paragraph.

<sup>9</sup> Rousseau 1762 [1964]: i.vi, last paragraph.

<sup>10</sup> See, for example, Isaiah Berlin, “Two Concepts of Liberty,” in his 2002: 166–219.

Turning to more recent times, Kantian autonomy is obviously not the same as the freedom affirmed by Jean-Paul Sartre and other existentialists, who rejected both causal determinants *and rational constraints* on our choices (see, for example, Sartre 1956 and 1948). To think that one *must* do something because laws of *reason* require it, according to this existentialist view, is just *bad faith*. One inevitably chooses one's way of life, affirming or revising it at each moment, in the absence of any divine, political, or even internal authority that determines what one ought to do. Sartre apparently accepted Kant's idea of negative freedom and his critique of all previous moral theories but rejected Kant's claims that a 'lawless' free will is impossible and that a free will must be subject to universal rational principles.

R. Hare's theory of universal prescriptivism is similar in some respects to Sartrean existentialism, but more qualified and subtle in ways that might suggest (mistakenly) that Hare's theory simply re-expresses Kant's universal law formulation of the Categorical Imperative.<sup>11</sup> The act-guiding principles that we endorse in our moral judgments and decisions, according to Hare, commit us to universal principles requiring or permitting anyone in relevantly similar circumstances to act in a relevantly similar way. Reason requires us to be consistent and clear about the implications of our choices, and Hare argued that this is more restrictive than one might think. In *Moral Thinking* Hare argued that, given human nature and circumstances, we can expect that virtually no one who is properly informed and thinking clearly would endorse universal prescriptions contrary to Hare's (the 'archangel') version of rule utilitarianism. In the end, however, his theory allows that there is nothing necessarily irrational or immoral about acting on principles commonly thought to be immoral provided one is *in fact willing* to live with their universal implications when applied to and by others. Hare draws from Kant's idea that a will with autonomy is 'a law to itself,' in a sense, but he ignores or rejects the requirements of Kantian rational autonomy that go *beyond consistency among our individual desire-based universal prescriptions*.

More recently feminist critiques of social contract theory and of psychological theories of moral development have again made various notions of autonomy the center of controversies. Carol Gilligan's critique of Lawrence Kohlberg's moral psychology, for example, raised

<sup>11</sup> R. M. Hare developed his universal prescriptivism in several works, *The Language of Morals* (Hare 1952), *Freedom and Reason* (Hare 1963), and *Moral Thinking* (Hare 1981).

philosophical doubts about the importance of autonomy for moral and political theory (see Gilligan 1982). Gilligan contrasted two ways of approaching moral questions. Her evidence seemed to show that an autonomy perspective was more prevalent in young men and a caring perspective more prevalent in young women, and this suggested to some that the high esteem for autonomy common in traditional moral and political theory has been due to male bias. From the autonomy perspective, according to Gilligan, moral problems are seen as all-or-nothing judgments, the options are accepted as given, and definitive rational solutions are expected (as in mathematics). From the perspective of care, moral problems are seen as embedded in a complex network of personal relationships so that ambivalence is warranted, the given options are not accepted as fixed, and definitive rational solutions are rarely to be expected.

Gilligan focused on everyday ('street-level') questions about how to confront difficult moral situations rather than the foundational questions that Kant addressed with his abstract conception of rational autonomy. (These everyday questions are apparently also the primary, if not exclusive concern in later developments of the "ethics of care," cf. Noddings 1984). Kant's thesis was not that we must confront particular moral decisions with what Gilligan describes as the autonomy perspective but rather that the *authority* of common-sense morality – what makes it binding – ultimately presupposes that we are (imperfect) rational agents with (Kantian) autonomy as a property of the will. Even though Kant suggests that certain substantive prohibitions (for example, of lying, suicide, and 'unnatural' sex) are absolute principles, these cannot be plausibly defended as implications of the fundamental moral principles that are based on the core idea of autonomy of the will. These fundamental principles, expressed as formulations of the Categorical Imperative, require consistency in willing, proper regard for humanity, and respect for (more specific) principles that can be ultimately justified from the point of view of everyone. These fundamental ideas do not entail that particular moral problems always present us with clear options, all-or-nothing choices, or definitive solutions that reason can determine without need for judgment and experience (cf. *GMS* 4:389). Kantian autonomy does not imply that we are, or should try to be, isolated individuals who do not depend on others. Moreover, the principles that rational autonomous persons would jointly endorse, given natural human circumstances and values, would have to acknowledge the importance of personal relationships (cf. *TL* 6:469–74).



#### 4 COMMON CONCEPTIONS OF AUTONOMY IN PRACTICAL ETHICS AS DISTINCT FROM KANT'S

There are, of course, many other noteworthy philosophical discussions of autonomy.<sup>12</sup> For example, various ideas of autonomy appear in discussions of practical questions ('applied ethics').<sup>13</sup> These are too many and too various to review in detail here, but in this section I will sketch three commonly recurring broad themes. Once again, these are conceptions of autonomy that must be distinguished from the abstract Kantian conception. In the final section, however, I will propose that values inherent in the Kantian conception can provide a deliberative framework for supporting the three common conceptions and making them more specific.

Common conceptions of autonomy in contemporary discussions, I think, include the following. *First*, autonomy is a *right* to make one's own decisions, especially about matters deeply affecting one's own life, without undue interference by others. Such a right must be qualified because each person's decisions inevitably affect others. The scope of the right, what counts as 'one's own decision,' and the criteria for 'undue interference' need to be specified. *Second*, autonomy is seen as a *capacity and disposition* to make decisions with *due reflection and independence* of mind. This is regarded as valuable to those who have it, a capacity they have reason to develop and exercise and others have reason to respect. Again, what counts as appropriate reflection and independent thinking needs to be specified. *Third*, autonomy is seen as being in control of a life that encompasses a substantial range of activities and relations with others. To live as an autonomous agent, in this sense, one needs to have a significant area of life to govern, not merely to direct one's inner thoughts but to make choices that effectively direct the course of one's life.<sup>14</sup> Having autonomy in this sense is valued beyond having a right that others not interfere with one's choices. It is not merely making one's own decisions among severely limited options, for example, to labor in the coal-mines or starve, to be subservient to a husband or forgo having children. Once again, however, the idea of autonomy needs to be filled out. There may be different ways

<sup>12</sup> For example, J. S. Mill, in *On Liberty*, ch. 3, defends the value of individuality, independence of mind, non-conformist 'experiments in living,' which are often associated with autonomy. See also Christman 1989; Dworkin 1989; and Mele 1995.

<sup>13</sup> For example, Beauchamp and Childress 2008: ch. 4.

<sup>14</sup> This ideal includes but goes beyond a right of autonomy and valuing capacities of autonomy, as I explain in "Autonomy and Benevolent Lies," in Hill 1991: especially 35–37.

of specifying the appropriate kind and area of control and the relevant conceptions of being in charge of one's own life.

These conceptions of autonomy for individuals mirror common thoughts about the autonomy of nation-states. *First*, they have an international right to make their own decisions within a certain jurisdiction without undue interference by other states. Autonomous states are not puppet states or colonies that are permitted a limited self-rule by another, more powerful state. *Second*, autonomous states are able to make more or less competent decisions from their internal deliberative processes, whether these are vested in a monarch, a ruling group, or the citizens. A state would fail to be fully autonomous (*self-governing*) if ruled by a wildly inconsistent madman, torn by civil war, or paralyzed by party politics. *Third*, autonomous states are thought to have a substantial area and range of activities to control or regulate. Although they are limited by territorial boundaries, they are effectively in charge of many important matters within their jurisdiction. A state is not autonomous if its 'government' is utterly ineffectual or has extremely limited powers to govern.

These three ideas associated with autonomy are not necessarily incompatible. At least it seems possible that for individuals more specific principles could largely resolve the potential conflicts that might arise between the different normative presumptions ('grounds of duty') implicit in the three general ideas of autonomy sketched above. If so, under a consistent and appropriated specified system of principles we could (without strict 'conflicts of duty') respect a right of autonomy for all persons, value their capacity and disposition for autonomy, and promote an ideal of their living effectively as autonomous persons. If the general presumptions can be harmonized in this way, we might think of the right, the value, and the ideal as aspects of one comprehensive conception of autonomy.

In any case, it is important to distinguish these practical conceptions (or aspects) of autonomy from Kantian autonomy. In Kant's foundational project a priori arguments are supposed to establish that autonomy, as an abstractly conceived property, must necessarily be attributed to the will of every moral agent because it is a precondition of moral obligation. We cannot discern empirically that some individuals have more or less than others, as is the case with intelligence, independence of mind, and effective ability to control events as one chooses.<sup>15</sup> Kantian autonomy is also

<sup>15</sup> In Kant's basic theory, I am assuming, autonomy of the will is an all-or-nothing or threshold concept. That is, although all human beings are imperfectly rational and individuals obviously differ in the degree to which their rational capacities have been developed and exercised, Kant thinks we must presume that in order to ascribe moral obligation and responsibility to

not a *right* but rather a feature of moral agency presupposed by all duties of right and virtue. Autonomy, then, is distinct from all of the common conceptions identified earlier – a moral right to make one’s own decisions, a capacity and disposition (in various degrees) to think and decide for oneself, and some significant degree of power to control events within one’s life.

Nevertheless, Kantian autonomy is not completely unrelated to these common themes. To the contrary, I propose in the next section that we can understand Kantian autonomy a possible source of the common ideas and a key to how they can be made more specific. That is, the Kantian idea of autonomy of the will is at the core of a framework for moral deliberation from which we can justify and specify more fully the common normative ideas of a right, a valued reflective capacity, and a power of effective control that are implicit in contemporary appeals to autonomy.

## 5 FROM KANTIAN AUTONOMY TO PRACTICAL VALUES AND PRINCIPLES

Having noted that Kantian autonomy is distinct from common practical conceptions, the question remains whether the gap between these can be reduced. More specifically, are there normative presumptions within the Kantian conception of autonomy that can serve as the basis for justifying and specifying practical principles corresponding to the three common conceptions identified in the last section? An affirmative answer, I think, is suggested by reviewing features of Kantian autonomy, though my discussion here must remain brief and speculative.

*First*, to begin with Kant’s conclusion, rational agents with autonomy necessarily have the capacity and rational disposition to respect and follow the Categorical Imperative as their supreme rationally authoritative standard, even if they never articulate it explicitly and often make choices that fail to live up to it. The main elements of Kant’s several formulations of the Categorical Imperative come together in his idea of a kingdom of ends, from which we can develop a Kantian framework for thinking

anyone we must suppose that the person has sufficient rational capacity and disposition to be self-governing by moral laws not based on inclinations. In Kant’s view, I take it, infants do not yet have Kantian autonomy and most mature adults do, but it is not clear how we are to judge whether it makes sense to apply the Kantian model of moral agency to human beings in a gray area between these. Mature moral agents may *express* their rational autonomy in action to various degrees, but Kant’s repeated insistence that we cannot judge their inner motives with assurance implies that we also remain uncertain about the degree to which they express their autonomy.

about specific practical principles.<sup>16</sup> A kingdom of ends is a systematic union of rational agents under common laws. The members of this ideal moral commonwealth are both authors of the (moral) laws and subject to them. They 'legislate' universal principles as rational persons with autonomy of the will and mutual recognition of their status as ends in themselves. Although we think of the members as having their own personal ends, we abstract from personal differences when considering what principles they would legislate.

The main point to draw from all this is that deliberation about specific moral principles aims to find principles acceptable to everyone who considers them rationally, setting aside particular personal interests, with due respect for humanity in each person. The next question then is whether deliberation from this point of view supports and helps to specify principles that correspond to the common practical conceptions of autonomy. As a step towards answering this question we must ask ourselves what values, or good reasons, would move the legislating members of a Kantian ideal moral commonwealth to adopt certain principles and reject others? A partial answer, I suggest, is already implicit in the thought that the members are rational and have autonomy of the will. In other words, our conception of ourselves and other moral agents as rational and autonomous commits us to certain values that can help to shape our deliberations about specific moral principles. For example, beyond the general respect for a shared moral perspective represented by the Categorical Imperative, the following further presumptions seem to be inherent in Kant's conception of moral agency.

*Second*, we have reason to regard every moral agent as having the capacities and dispositions of rational autonomy and every potential moral agent as having these in latent form. There should be a presumption, then, that human beings are not *mere* products of their history and culture, that they *can* think consistently and coherently about practical matters, and that they can recognize and respond to moral arguments. Although they may act selfishly, cruelly, or callously, we must be reluctant to dismiss them as too dense to understand or too amoral to care about whether their acts are offensive from the common point of view of humanity. At least we should assume that human beings have the capacities and dispositions of rational autonomy unless it becomes fully evident

<sup>16</sup> *GMS* 4:433–40. The Kantian framework for deliberation about moral principles sketched here, an extension and modification of Kant's explicit views, is discussed in Hill 2000: chs. 2, 4, and 8; Hill 2002: ch. 3; and Hill 2012: chs. 3, 8–11, and 13.

that it is impossible to make sense of them in this way. We cannot reasonably attribute rational autonomy to trees, insects, and the most severely brain-damaged human beings because we cannot make any sense of their current or expected behavior with those concepts.<sup>17</sup> There may be difficult borderline cases, of course, but insofar as we hold people to be morally responsible (or, in the case of young children, potentially so) we must regard them as (at least potentially) capable of responding to moral reasons. From our perspective it may seem that Kant too readily dismissed empirical evidence that some human beings ('sociopaths') lack the capacities and dispositions of rational autonomy that are necessary (in his view) for moral obligation and responsibility. His conceptual argument that rational autonomy is necessary for moral agency, however, is independent of his (and our) beliefs about how many and which particular flesh-and-blood human creatures are (at least potentially) moral agents.

*Third*, all persons with Kantian autonomy have a rationally motivating disposition to govern their own lives accordingly. Practical reason, in Kant's view, is not simply a capacity to understand abstract rational principles but a disposition to acknowledge their authority and to follow them. Insofar as we have rational autonomy, then, we deeply care about preserving, exercising, and developing our ability to live with others under principles shaped by the conception of persons as rational and autonomous.

This is a rational concern not merely to use one's reason in making immediate choices but also to take the necessary steps to protect and improve one's capacities for rational autonomous choice under challenging empirical conditions. None of this is possible in isolation. We need cooperative rational interactions with others for intrinsic as well as instrumental reasons. Kantian principles of rational autonomy, even important duties to oneself, are about how to govern oneself as one among many in a shared world.<sup>18</sup> We need rational exchanges with others to exercise as well as strengthen our capacities as rational autonomous persons.<sup>19</sup> There are presumptive reasons, then, to establish and maintain social practices and legal institutions that protect and support our ability to live with others

<sup>17</sup> Kant offers us a model of inter-acting rational moral agents in which there are necessary (a priori) relations among their capacities, dispositions, principles, obligations, and responsibilities, assuming certain general features of the human condition. These connections are not established empirically, but it must remain partly an empirical question whether the model fits – or can enable us to make sense of – the actual and expected behavior of various creatures, including members of the species *Homo sapiens*.

<sup>18</sup> Kant's "duties to oneself" include duties not to lie, to avoid servility, and "to be a useful member of the world" (*TL* 6:429–32, 434–37, 445–46).

<sup>19</sup> Andrews Reath develops this theme in "Legislating for a Realm of Ends," in Reath 2006: ch. 6, 17ff.

under mutually acceptable rational principles, and our ideal as individuals should be, as far as possible, to govern our own lives by such principles.

These three basic ideas, which are inherent in Kant's attribution of rational autonomy to all moral agents, offer a basis for developing more specific principles defining a *right* of autonomy, appropriate respect for a *valuable capacity* for rational autonomy, and an *ideal* of living fully as a rational autonomous person. The details must partly depend on the existing background conditions in the context for which principles are intended – the culture, institutions, social practices, and economic structures that can pose threats to autonomy but can also be made to affirm it. Whatever the details may be for various contexts, we can expect the principles to reflect the three kinds of normative presumptions we have been considering.

To conclude, then, let us return briefly to the three kinds of autonomy considerations that I sketched previously as commonly found in discussions of practical issues.

*First*, is there a moral *right* of autonomy? In developing ('legislating') specific moral principles from the common point of view ('kingdom of ends') we would be rationally motivated to preserve, develop, and exercise our capacity for rational autonomous living under shared principles equally acceptable to others. In the imperfect actual human circumstances, one person's choices often interfere with others' even when they are all well-meaning. Malicious, cruel, and callous people pose further threats to our ability to live autonomously. Our range of choices can be illegitimately restricted in many ways – by physical force, coercive threats, deception, manipulation, and oppressive ideologies that damage our capacities for rational self-governance. Given these circumstances, we have strong reason (as rational moral 'legislators') to establish and maintain principles that affirm each person's right to govern his or her own life within certain acknowledged limits. What these limits are depends on what is needed to counteract illegitimate interferences with the range and kind of control that each would have under principles acceptable to all. As Kant argued, a legal system is needed not only to define and adjudicate rights specifically but to use coercive force to "hinder hindrances to freedom" (*RL* 6:230–33). Arguably, however, moral principles would affirm a broader moral right of autonomy for individuals and social practices apart from legal enforcement, because lies, manipulation, and oppression can severely damage our capacities and restrict our opportunities for rational autonomy in ways that even just legal systems cannot prevent. Local social norms typically define and restrict such autonomy rights within different

cultures, and the Kantian legislative framework can serve as a basis for criticism and reform of these norms.

*Second*, are there presumptive reasons, beyond rights, to *value each person's capacity for rational autonomy*? For example, should it be an aim of education and self-improvement to develop the ability, as Kant urges, to think for oneself about practical issues, consistently and from the point of view of everyone (cf. *KpV* 5:29–35)? Again, details will justifiably vary with local conditions, but insofar as we conceive of persons as rational and autonomous we are committed to regarding their capacity for rational autonomy as a value we must respect. By hypothesis all Kantian legislators have a deep rational disposition to exercise, and so to preserve and develop their own capacities for rational autonomy, and their commitment to shared reciprocal principles requires them to respect the same disposition in others. In practice this means not only protecting, developing, and exercising one's own capacities for rational autonomy within limits that appropriately respect others', but also developing and supporting social institutions and relations that promote the same value for everyone, for example, through schools, community organizations, standards for ethical journalism, prison reform, and so on. Although students, as well as supposedly mature adults, too often remain prone to blind conformity, lazy indifference, or thoughtless rebellion, educational institutions of various kinds could try to develop more effectively the valuable ability to reflect rationally and reasonably about personal, social, and political decisions.

*Third*, are there good presumptive reasons to find and establish principles that promote the opportunities of each person to live a full life as a rational autonomous person? Once again, the Kantian position considered here strongly affirms the general presumption but acknowledges specific concerns about the possible means. We have reason to care about each person's opportunities and resources for living effectively in control of his or her own life, but we can promote this value only under principles that are fair to all and respect moral constraints on the permissible means. What specific principles and permissible means are appropriate for various contexts depends in part on the circumstances. Trying to achieve utopia by means of murderous force is not an option, but within a range of acceptable possibilities arguably those most affected must work out the specific principles, institutions, and practices that best respect and promote the ideal of autonomous living for all in their context.

Some contemporary readers may regard the conclusion here as too idealistic and too tied to abstract Kantian terminology, but I think that

essentially the same points can be re-expressed in more familiar terms and for the most part recognized, not as what now is, but as what ought to be reflected in practice. Others may be disappointed that Kantian autonomy by itself does not give us more definite answers to particular questions, for example, about informed consent, paternalistic laws, and respect for the mentally impaired. One should not, however, expect too much – or things of the wrong kind – from general moral theory, which at its best seeks a deeper understanding of our moral beliefs and their grounds but is not a substitute for hard thinking and conscientious joint effort to resolve specific practical problems.