KANT ON MORAL AUTONOMY

EDITED BY
OLIVER SENSEN
Tulane University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9781107004863

First published 2013

Printed and Bound in the United Kingdom by the MPG Books Group

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Kant on moral autonomy / edited by Oliver Sensen.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-107-00486-3 (hardback)
1. Kant, Immanuel, 1724–1804. 2. Free will and determinism.
I. Sensen, Oliver.
B2799.F8K36 2013
170.92–dc23 2012020416

ISBN 978-1-107-00486-3 Hardback

In honor of Onora O'Neill
Admired colleague, friend, and mentor

CHAPTER 2

Kant's conception of autonomy of the will

Andrews Reath

1 INTRODUCTION

Kant tells us that "autonomy of the will is the property of the will by which it is a law to itself (independently of any property of the objects of volition)" (*Autonomie des Willens ist die Beschaffenheit des Willens dadurch derselbe ihm selbst [unabhängig von aller Beschaffenheit der Gegenstände des Wollens] ein Gesetz ist*) (*GMS* 4:440). He goes on to say that the law that the will is to itself (the "principle of autonomy") is a version of what has been identified as the basic principle of morality: "to choose only in such a way that the maxims of your choice [*Wahl*] are also included as universal law in the same volition." It is rare for Kant to define an idea of such import so succinctly and straightforwardly, but that does not mean that his conception of autonomy is readily understood. There is much to unpack in these claims, but this much is quite clear: autonomy is a property of the will – i.e., of the faculty of rational volition, which Kant elsewhere identifies with practical reason. (See *GMS* 4:412–13 and *RL* 6:213.) This property is that the will is a law to itself ("independently of any property of the objects of volition" [*GMS* 4:440]). Finally, the law that the will is to itself is the Categorical Imperative. Let's call this cluster of ideas 'Kant's thesis of autonomy of the will.' A standing problem is how to reconcile Kant's thesis of autonomy with the objectivity of moral principle to which he is committed. Given the necessity and universality of the moral law, in what sense is the rational will a law to itself, or does it 'give itself' its own fundamental law?

In a series of essays I have argued that Kant's thesis of autonomy of the will should be interpreted somewhat narrowly as the sovereignty of the will over itself, which I understand to have both a negative and a positive dimension (Reath 2006: chs. 5–7 and Reath 2008). Negatively the rational will is not bound by any externally given principles or authority. Positively it is a law to itself in the sense that (I argue) the nature of

rational volition is the source of its own fundamental norm, a principle that Kant identifies as the Categorical Imperative. Referring to the will's 'sovereignty' here acknowledges the political notion of autonomy as the model for Kant's conception of autonomy of the will.[1] A political sovereign does not owe obedience to any outside authority, and it has a lawgiving power established by its constitution or *loix fondamentales*.[2] Just so, the rational will is a sovereign lawgiving power. It is in some sense the source of the fundamental normative principles that govern its activity, and is thus subject "only to *its own* though still *universal lawgiving*" (*GMS* 4:432). In this essay I offer an overview of my attempts to spell out the precise sense of these large claims and add some further details to this interpretive approach.

Kantian autonomy is often understood as the capacity to be motivated by principles of pure practical reason, independently of inclination or empirically given interests. This reading makes autonomy a recognitional *cum* motivational capacity – for example, the capacity to accept or impose on oneself principles of pure practical reason out of recognition of their authority and to determine oneself to act from these principles independently of empirically given incentives and interests.[3] Or given Kant's tendency to equate autonomy with freedom of the will, it is common to think that his robust notion of transcendentally free agency is the central element of his conception of autonomy.[4] However, I find that Kant's canonical remarks suggest a focused and more confined notion of autonomy as, for lack of a better term, a normative property or normative standing of the rational will, which applies derivatively to individual rational agents: the autonomy of the will is its normative independence (that it need not answer to any outside authority, but only to itself) and the fact that it is the source of its own fundamental norm – the moral law. That makes it normatively self-governing. And as we shall see, far from there being any inconsistency between autonomy of the will and the necessity of the moral law, Kant thinks that the necessary authority of moral principle requires and is explained by the thesis of autonomy. I suggest that in order to understand Kant's conception of autonomy, we do best to begin from this confined notion and build out to related ideas,

[1] The political origin of the concept of autonomy is often noted. See Schneewind 1998: 3n. 2, 483; O'Neill 2000: 40; and Darwall 2006a: 263.

[2] Cf. Jean-Jacques Rousseau, *Du contrat social* (1762 [1964]: II.xii; also I.v–vi).

[3] Note that a rational intuitionist can agree that we have this complex capacity. But intuitionism holds that we are responsive to externally given objective laws and would not accept Kant's theses about autonomy of the will.

[4] See, e.g., *GMS* 4:447, 454 and *KpV* 5:33.

such as the motivational capacity to act from rational principle independently of sensibly given interests, the freedom of will, the moral standing of individual rational agents as ends in themselves, and so on.

In the next section, I survey the passages in the *Groundwork* in which Kant lays out his thesis of autonomy of the will. I argue that this thesis is best approached through the idea that there are internal formal principles of rational volition that, as it were, specify its nature. If this is right, investigating Kant's thesis of autonomy requires us to venture into larger issues about the normative authority of reason. The thesis of autonomy of the will is a special case of the more general idea that reason has autonomy and is a self-determining faculty. Kant's claim that the will is a law to itself (as I interpret it) presupposes a certain conception of the will, and in section 3 I take that up. Finally in section 4, I show how to build out from the confined normative reading of autonomy of the will to other related ideas.

2 AUTONOMY OF THE WILL: A SKELETAL ACCOUNT

Two sets of passages in the *Groundwork* are the source of Kant's conception of autonomy of the will. The first is the initial appearance of the idea of autonomy – "the idea of the will of every rational being as a will giving universal law" (*als eines allgemein gesetzgebendes Willens*) – at *Groundwork* 4:431, where Kant makes the following claim:

> Hence the will [*der Wille*] is not merely subject to the law but subject to it in such a way that it must be viewed also *as giving the law to itself* [*als selbstgesetzgebend*] and just because of this as first subject to the law (of which it can regard itself as the author [*Urheber*]). (*GMS* 4:431)

It is a few paragraphs later that Kant claims

> that the human being [*der Mensch*] ... is subject only to *his own* but still *universal* lawgiving [*seiner eigenen und dennoch allgemeinen Gestetzgebung*] and that he is bound to act only in conformity with his own will, which, however, in accordance with nature's end, is a will giving universal law. (*GMS* 4:432)[5]

The second passage is the claim from *Groundwork* 440 with which we began, and is repeated at the opening of the third section (*GMS* 4:447). These passages focus initially on 'the rational will,' which is most naturally understood as a faculty or rational capacity found in individuals, though

[5] In the paragraphs following *Groundwork* 431 Kant refers several times to the "will of every rational being," and sometimes to every human will, as "*ein allgemein gesetzgebenden Wille.*"

Kant also gravitates towards referring to individuals – for example, "this lawgiving must ... be found in every rational being himself" (*GMS* 4:434), or dignity is found "in the person who fulfills all his duties ... insofar as he is at the same time lawgiving"[6] (*GMS* 4:440). I'll discuss both passages, beginning with the second, which I take to state the central idea.

To get a handle on the sense in which the will is a law to itself, one should note that the thesis of autonomy is introduced to explain the authority ascribed to moral principles in ordinary moral thought – that moral requirements are taken to apply unconditionally and to have deliberative priority over other kinds of reasons.[7] At *Groundwork* 431 and following, Kant is arguing that it follows analytically from the concept of a categorical imperative that a will or an agent subject to such a principle must be regarded as "giving the law to itself" [*als selbstgesetzgebend*]. A practical principle that does not "arise from" the will of the subject [*aus seinem Willen entsprang*] "has to carry with it some interest by way of attraction or constraint," in which case the principle is a hypothetical rather than a categorical imperative (*GMS* 4:432–33). Likewise, after claiming at *Groundwork* 440 that the thesis of autonomy of the will can be shown "by mere analysis of the concepts of morality," Kant explains why moral theories of heteronomy fail. The theoretical alternative to basing morality in a principle that the will gives to itself is to base it in a "property of an object of volition" – that is, to base it in some feature of a potential object of volition in which we are assumed to have an interest. This is the method in moral philosophy of beginning with the concept of the good without prior investigation of the nature or a priori commitments of volition, the method that Kant rejects in the *Critique of Practical Reason* (*KpV* 5:63–65). Heteronomy always results because "the object, by means of its relation to the will, gives the law to it" (*GMS* 4:441). And the problem with such moral theories is that they represent moral principles

[6] Note that if the will is practical reason, then it follows that practical reason has autonomy and is a law to itself. Kant tends to refer to the autonomy of practical reason in the *Critique of Practical Reason*; see, e.g., *KpV* 5:33, 43. The claim that practical reason is a law to itself does not lend itself to the same individualistic or voluntaristic reading that Kant's claims about the autonomy of the will are often thought (incorrectly) to suggest.

[7] One qualification: since the thesis of autonomy of the will is part of the analytic argument of *Groundwork* II, it does not establish the authority of morality for us (nor that we are legislators of moral principle). Kant's analytic claim is that any agents who are indeed subject to the moral law must be regarded as legislating and that their legislative role will figure in the explanation of its authority. To establish that we are such agents, Kant needs to appeal to some (synthetic) fact about our volitional capacities or practical self-consciousness. Paragraph 57 makes it clear that the thesis of autonomy follows from the necessity that is a component of our ordinary conception of moral requirement. For further discussion, see Reath 2006: 99–101, 134–42.

as hypothetical imperatives whose normative force is conditional on having an interest in that object or substantive value (without establishing that this interest is rationally necessary).[8] Thus Kant accepts the striking claim that the unconditional authority or categorical nature of moral principle is genuine only if the moral law arises from the will or is a law that the will gives to itself. Why is this?

A general link between normative necessity and autonomy is forged by the idea of a formal principle, by which I mean the internal constitutive norm(s) of a form of cognition or rational activity. A formal principle is a principle based in the self-understanding of a kind of rational activity that defines and makes it possible to engage in that activity and that, accordingly, tacitly guides all its instances (even those that are defective or incomplete). As such it is uniquely suited to govern the activity with necessity because one cannot coherently reject the principle while engaging in, or conceiving oneself to engage in that activity.[9] But as the internal principle based in the self-understanding of a kind of rational activity, a formal principle specifies and 'arises from' the nature of that activity.

The central insight that drives the arguments of both the *Groundwork* and the second *Critique* is that to establish the authority of the morality, one must show that its basic principle, the Categorical Imperative, is the formal principle of rational volition – the internal principle based in the self-understanding of rational volition through which one exercises that capacity. The formal principle of rational volition would be the internal norm that specifies and arises from the nature of rational volition, and in that sense is the law that the will is to itself. That is to say that the normative authority of moral requirement is genuine only if based in a principle that arises from the nature of the rational will, in the law that the will is to itself – independently of any contingent interest in the properties of potential objects of volition.

This line of thought is at work at different points in the *Groundwork* and the *Critique of Practical Reason*. In both *Groundwork* II and the

[8] One of Kant's points in the second *Critique* is that moral theories of heteronomy lack the resources to show that this interest is rationally necessary. They try to derive practical principles by presenting some object as intrinsically good, or as having features that are reason-giving. But if a theory begins from a claim about the good without prior investigation of the a priori features of volition, the only available standard of good is whether the representation of the object produces interest. Since that is a matter of feeling and thus empirical, this method cannot identify genuine practical laws. For further discussion of this point, see Reath 2010: 49–51.

[9] Here I draw on Reath 2010: 41–48. See also Korsgaard 2008: 7–10, who stresses that constitutive norms are both descriptive and normative.

second *Critique*, Kant derives a statement of the moral law from an analysis of practical reason as a faculty of principles. Roughly, the idea of the complete determination of action through reason leads to the idea of a practical law, and from the idea of a practical law or the concept of a categorical imperative, Kant derives a statement of the Formula of Universal Law (FUL) (*GMS* 4:412–21 and *KpV* 5:19–30). In similar fashion, he derives a statement of the Formula of Humanity from analysis of practical reason as a faculty of ends (*GMS* 4:427–29).[10] Setting aside the details of these arguments, what is important for our purposes is that Kant's method of argument here indicates his belief that analysis of the nature of practical reason leads to a statement of its fundamental norm – that practical reason is a law to itself.

Furthermore, a central component of the foundational arguments of each work is the claim that the moral law (FUL) is the basic principle of a free will. *Groundwork* III tries to establish the authority of the moral law by arguing (1) that "a free will and a will under moral laws are one and the same" (*GMS* 4:447) and (2) that it is a necessary feature of practical self-consciousness that we act under the idea of freedom and that we identify with our capacity for free agency as "our proper self" (*GMS* 4:448, 458 and 461). The first is best understood as the claim that the moral law is the formal principle of free agency and in that sense the law that the free rational will is to itself.

In sum, the thesis of autonomy of the will is introduced to ground the necessary authority of moral principle. Kant tries to establish the necessity of the moral law by arguing that it is the formal principle of rational volition – the internal constitutive principle that arises from its nature and specifies what it is to exercise the will. This idea gives the sense in which the will is a law to itself (independently of any property of the objects of volition) and it is the fundamental idea underlying the thesis of autonomy of the will. Here it is worth noting that, while Kant sometimes says that the will "gives" itself a law (*GMS* 4:441), his canonical phrasing is that the will *is* a law to itself. There is no thought here that this law is 'given' or 'legislated' by a particular act of volition, i.e., through a discretionary or voluntary act of individual moral subjects. The thesis of autonomy is rather that the nature of rational volition generates or supplies its own fundamental norm.[11]

[10] I develop this reading in Reath in press.

[11] For discussion see Reath 2006: 104–9.

While autonomy is principally a property of the will as a rational faculty, the passages following *Groundwork* 431 suggest that individual moral agents must be regarded as legislating or giving universal law and are subject only to their own will or (universal) lawgiving. Certainly if the will is a law to itself, it follows that individual agents are unconditionally bound only to this fundamental law – the law that, as it were, confers the capacity for rational volition and makes us agents – because that is the only rational principle with genuine normative authority for agents qua rational. But with the above conception of autonomy in place, we can identify a derivative sense in which individual agents must be regarded as legislating the particular substantive moral requirements to which they are subject. What is needed here is that the Categorical Imperative (FUL) is the basis of a form of practical reasoning that identifies some substantive moral requirements – a form of reasoning whose authority comes from the fact that it is the 'constitution' of the rational will along the above lines (cf. Reath 2006: 109ff., 129ff.; Reath 2008: 129–30). We may also need this reasoning to 'generate' moral requirements – for example, by applying the formal condition of universal validity that comes from the nature of volition to the necessary interests and ends of rational agency (such as the interest in exercising one's agency, in directing one's will through one's own practical judgments, etc.). One might then mount the following argument from the concept of an unconditional requirement on action (suggested by *GMS* 431–32): Since a categorical imperative applies unconditionally, its normative authority does not come from any contingent interest in an agent, but must come from the reasoning that makes it a law. That is, sufficient reason to comply with the principle is given by the reasoning that shows that the principle has the form of a law. But an agent who is bound simply by the reasoning that makes a principle a law must have the capacity to carry that reasoning out, and – noting that a legislator is an agent with the power to give law through his will by carrying out an authoritative legislative procedure – thus has the same basic capacities as would be required of its legislator. In that sense, agents subject to practical laws 'must be regarded as legislating.'[12]

[12] For attempts to work this idea out, see Reath 2006: 99–104, 137–42 and Reath 2008: 128–29. In the end it is hard to make the case that individual agents 'give particular laws through their willing' and the best one can hope for here is a meaningful analogy (as suggested by the phrasing that moral agents "must be regarded" [*angesehen werden muß*] as legislating). But I now think that this sense of autonomy is less central than the idea that the will is a law to itself as construed above.

Do the above senses of autonomy support the claim that either the rational will or the individual moral agent is (or can meaningfully be regarded itself as) the "author" (*Urheber*) of moral principle – a phrase that Kant uses at *Groundwork* 431 (the will "can regard itself as the author" of moral law)? Kant distinguishes between the "author of a law" and the "author of the obligation in accordance with a law" in the following (now often cited) passage:

> One who commands (*imperans*) through a law is the *lawgiver* (*legislator*). He is the author (*autor*) of the obligation in accordance with the law [*Urheber der Verbindlichkeit nach dem Gesetz*], but not always the author of the law [*Urheber des Gesetz*]. In the latter case the law would be a positive (contingent) and chosen [*Willkürlich*] law. A law that binds us a priori and unconditionally through our own reason can also be expressed as proceeding from the will of a supreme lawgiver, that is, one who has only rights and no duties (hence from the divine will); but this signifies only the idea of a moral being whose will is a law for everyone, without his being thought as the author of the law. (*RL* 6:227; cf. *Collins* 27:282–83)

The "author of a law" is the agent who, as it were, writes the law or determines its content at his discretion, while the "author of the obligation in accordance with the law" is an agent whose will makes a law binding – for example, an agent with authority over some group of agents who, by addressing a law to them, gives them a reason of authority to comply. Kant claims that the only laws authored in the first sense are those whose content is contingent, such as positive social law.

Various commentators have rightly drawn our attention to Kant's claim that moral principles, as necessary principles of reason, have no author in this sense. However they overlook the fact that Kant allows that there can be an 'author of the obligation' of such laws and that by 'lawgiver' he often means the latter.[13] Since 'moral obligation' though not moral laws

[13] See Kain 2004 and Wood 2008: 111–14. I discuss this passage in Reath 2006: 145–49. See also the passage from the *Lectures on Ethics* [Collins]: "Anyone who declares that a law in conformity with his will obliges others to obey it, is giving law. The lawgiver is not always simultaneously an author [*Urheber*] of the law; he is only that if the laws are contingent. But if the laws are practically necessary, and he merely declares that they conform to his will, then he is a lawgiver. So nobody, not even the deity, is an author of moral laws, since they have not arisen from choice but are practically necessary … But moral laws can still be subject to a lawgiver; there may be a being who is omnipotent and has the power to execute these laws, and to declare that this moral law is at the same time a law of his will and obliges everyone to act accordingly. Such a being is then a lawgiver, though not an author" (*Collins* 27:282–83).

In both passages, God, though not the author of the moral law, can be conceived as the author of moral obligation and is a lawgiver in that sense. In a related passage (also cited by Wood) Kant writes: "Crusius believes that all obligation is related to the will of another. So in his view all obligation would be a necessitation *per arbitrium alterius*. It may indeed seem that in an

can have an 'author,' "author" at *Groundwork* 431 must refer to "author of the obligation." More generally, what is at issue in the thesis of autonomy and talk of the will as *allgemein gesetzgebend* – what these ideas are intended to ground – is the authority of moral principle, not its content. Thus there is no reason to deny that the rational will is 'author' of moral obligation in that its normative authority comes from the fact that it is based in the law that the will is to itself.

This skeletal account of Kant's conception of autonomy of the will raises certain questions. To claim that the will is a law to itself "independently of any property of the objects of volition" is to hold that there is an alternative to the method of heteronomy and that the nature and a priori commitments of rational volition suffice to provide its own fundamental norm. For that to be a live possibility, rational volition must have a nature – e.g., a formal normative aim implicit in all exercises of volition, as suggested by Kant's remark that the will is "in accordance with nature's end a will giving universal law" (*GMS* 4:432: *dem Naturzwecke nach aber allgemein gesetzgebenden Willen*). That is to say that Kant's thesis of autonomy presupposes a (not uncontroversial) conception of rational volition as having the formal aim of operating according to universally valid principles, or of willing the good.[14] A further problem is that it is not clear that this account has yet identified a genuine notion of 'legislation' or 'giving law,' if that requires a particular act of volition. Is the idea that the moral law is the formal internal norm of rational volition all there is to the idea that the will 'gives' itself the moral law? That is still a pretty substantial idea that supports the sovereignty of the will over itself – that it is normatively independent and the source of its own fundamental law. But it may seem to leave the idea of 'giving' law thin and somewhat obscure. I take on these questions in the next section by turning to Kant's conception of the will.

obligation we are necessitated *per arbitrium alterius*; but in fact I am necessitated by an *arbitrium internum*, not *externum*, and thus by the necessary condition of universal will; hence there is a universal obligation" (*Collins* 27:262).

It's worth noting that in this passage Kant does not reject the view that all obligation is related to will, but only that it is related to the will of another. One can be necessitated by *arbitrium internum*. Taken together, these passages show that Kant leaves room for an 'author of moral obligation,' where the authority of moral obligation comes from the rational will.

[14] Onora O'Neill has stressed that Kant's conception of autonomy (which she interprets as the capacity to act from principles whose justification comes solely from reason, principles that can be freely adopted by a plurality of agents) requires a non-empiricist view of practical reason and action. See O'Neill 2000: 41ff. and O'Neill 2003: 7ff. This is an important point that is often overlooked. I agree that Kant's conception of autonomy grows out of a particular conception of volition and action, and in the next section indicate what I think it is.

3 THE WILL

Since autonomy is a property of the will, some consideration of the will should bring certain details of Kant's conception of its autonomy into clearer focus. Kant's understanding of the will is no simple matter, but the main ideas are found in the following texts. The will is a rational causality – "a kind of causality of living beings insofar as they are rational" (*GMS* 4:446). In the *Groundwork* Kant initially defines the will as "the capacity to act from a representation of laws, that is from principles." Since acting from principles involves "the derivation of actions from laws," which requires reason, Kant identifies the will with practical reason (*GMS* 4:412). He implies that the will is the capacity to choose [*wählen*] what "reason independently of inclination cognizes [*erkennt*] as practically necessary" or good, where judgments of "practical good" are supported by "grounds that are valid for every rational being as such" (*GMS* 4:412; see also *KpV* 5:61).

In the *Metaphysics of Morals*, Kant defines the will as the faculty of desire in accordance with concepts "whose inner determining ground, hence even what pleases it, lies within the subject's reason" (*RL* 6:213). A faculty of desire is the capacity in a living being "to be, by means of its representations, the cause of the objects of these representations" (*RL* 6:211; cf. *KpV* 5:9n. and *KU* 5:220). Will is a faculty of desire in which the (desiderative) representations that guide the activity of the living being to realize their objects are based in reason – for example, they are rational principles, or practical judgments that represent an action or end as good by 'deriving' it from a principle. Further, this passage famously distinguishes *Wille* (will) and *Willkür* (choice) as different aspects of "the faculty of desire in accordance with concepts" (*RL* 6:213), which is the will or faculty of volition in a broad sense. The power of choice [*Willkür*] is the faculty of desire in accordance with concepts "insofar as it is joined with consciousness of the ability to bring about its object by one's action." In the narrower sense, the will [*Wille*] is "the faculty of desire considered not so much in relation to action (as the power of choice is) as rather in relation to the ground determining the power of choice to action, and has itself properly no determining ground, but is, insofar as it can determine the power of choice, practical reason itself" (*RL* 6:213).

These passages characterize the will (or practical reason) as the complex capacity in rational agents to move from rational principle to action through their own representational activity. It involves the capacity to form rational judgments that represent an action or end as good – judgments

that, as determinations of the 'faculty of *desire* in accordance with concepts,' carry motivational force – and the capacity to determine oneself through those practical judgments to realize the object represented therein as good. The tight connection between will and practical *reason* in these passages suggests that practical judgments derive action from (purportedly) universally valid principles and are judgments to the effect that an action or end is supported by good and sufficient reasons.

In his recent work, Stephen Engstrom has made it clear that Kant understands volition to have two interrelated moments, both a rational or cognitive and a desiderative or causal moment, and that self-consciousness plays an essential role in each moment (Engstrom 2009: 25–65 and Engstrom 2010: especially 44–46). Volition involves practical reasoning aimed at judgments representing an action or end as good that make a claim to universal validity. This rational or cognitive moment is self-conscious in that rational agents are conscious of themselves as subjects with the capacity for practical judgment, and in deliberation they understand themselves to be moving toward a practical judgment with a claim to universal validity (to be determining what they have good and sufficient reason to do). Practical judgments are also forms of desiring that strive to be efficacious in realizing their objects (Engstrom 2010: 44). The causal moment in volition is the exercise of the general capacity to determine oneself through one's practical judgments to realize the objects that they represent as good. The causality of volition is self-conscious because it is efficacious in realizing its objects through understanding itself to be efficacious. Self-consciousness of oneself as an agent with the general causal capacity to realize the objects of practical judgments thus figures in the representational activity that guides volition (Engstrom 2010: 32, 45–46).

Following Beck and Allison, Kant's use of *Wille* and *Willkür* is standardly interpreted to mark a distinction between "the legislative and executive functions of a unified faculty of volition, which [Kant] likewise refers to as *Wille*" (Allison 1990: 129; cf. Beck 1960: 199–202). Engstrom argues persuasively that the distinction between *Wille* (in the narrow sense) and *Willkür* should be traced to the distinction between the cognitive and the causal moments of volition, and to the self-consciousness that guides each. *Wille* (in the narrow sense) is the capacity underlying the cognitive moment of volition – a capacity for practical reasoning – and its exercise is guided by self-consciousness as a subject with the capacity for practical judgments of good (as well as principles and ends that can initiate practical reasoning).

Willkür is the capacity underlying the causal moment of volition. Its exercise is guided by one's general self-consciousness as an agent with causal powers, in conjunction with a representation of one's ends and information about one's actual causal capacities. Its "office," Engstrom writes, is to bring one's practical judgments into agreement with one's "empirically determined cognizance of oneself as agent, ensuring that ends set in acts of choice are … within one's capacity and that the actions chosen are sufficient for reaching them" (Engstrom 2010: 47). The idea, I take it, is that the exercise of *Willkür* results in the representation of an action within one's power that one believes will achieve an end and that this representation determines and guides one's powers of agency. (Here choice may involve adjusting one's ends to one's causal powers as well as finding suitable means.) The exercise of the power of choice issues in a judgment about action that, as it were, completes the derivation of an action from principles. This makes *Willkür* the executive aspect of volition in the very literal sense that its function is to settle on actions that carry out the judgments and realize the ends of *Wille*.[15]

If this interpretation is right, *Willkür* is not a purely elective capacity that decides whether or not to follow the conclusions of practical reasoning, as is often thought.[16] Rather, it is the causal or executive aspect of the complex capacity to move from rational principle to action via practical judgment. In order to exercise the capacity to realize the object of one's representations through those representations, there must be a

[15] Engstrom suggests that the basic 'act' of the faculty of volition is practical judgment and that both the will and the power of choice contribute to practical judgment (Engstrom 2010: 44, 46). His basic picture (somewhat simplified) is that the exercise of the will (in the narrow sense) leads to the judgment of an end as simply good. This judgment is a 'wish,' which is a desiderative state that abstracts from an agent's actual causal powers (cf. Engstrom 2009: 66–69). The wish for an end initiates further deliberation that goes into the exercise of the power of choice. This reasoning seeks to bring one's ends into agreement with one's causal powers by finding actions that will realize the end (means), or by adjusting the ends if necessary (thus re-invoking practical reasoning about ends). The outcome is the representation of an action within one's power that will realize the end, or more specifically, the judgment of such action as good. This judgment is the act of choice that completes the derivation of an action from principle and guides the deployment of an agent's causal powers.

[16] I believe that on Engstrom's reading, bad choice would not be due to *Willkür* setting out on its own and flouting the dictates of *Wille*, but rather to a defect in the cognitive moment of volition (though possibly one that is 'motivated') – for example, defective practical reasoning and judgment. I find Engstrom's rendition of *RL* 6:213 convincing, but other texts lend support to the view of *Willkür* as an elective capacity that, though not definable as the capacity to choose in opposition to practical reason, gives agents the ability either to conform to or act against practical reason (cf. *RL* 6:226). For example, the elective interpretation fits much of the *Religion*, in particular the idea that the propensity to evil "attaches to the moral faculty of choice [*dem moralischen Vermögen der Willkür*]" (*RGV* 6:31). To resolve this interpretive issue, one would have to address these and other passages, but that is beyond the scope of this chapter.

representation of some end on the table, and that is the purview of the cognitive moment of volition. So choice comes into play only when practical reasoning and judgment has put an end on the table. The charge of the power of choice is then to execute the practical judgment, not to choose whether or not to follow it – as though, once we reach a judgment about what is good or supported by reason, we must still decide whether or not to do it.[17]

Let me now draw out some of the implications of this conception of the will for Kant's thesis of its autonomy. The first concerns the conception of the will that is presupposed by the thesis that the will is a law to itself. Though I can't fully defend this claim here, I believe that Kant understands volition to be constitutively aimed at good.[18] At any rate, some such conception of volition is presupposed by his thesis of autonomy as I have interpreted it (that the nature of volition supplies its own fundamental norm). Roughly, volition has the formal aim of reasoning correctly from rational principle to practical judgments about good that carry an implicit claim to universality and that (as *practical*) can be the cause of the objects that they represent as good. That means that it is part of practical self-consciousness that rational agents understand themselves to have the capacity for practical judgment, and that in exercising their will they understand themselves to be aiming at practical judgments of good that carry an implicit claim to universal validity (that articulate good and sufficient reasons for action that others can also accept). These aims are necessary features of rational volition in that an agent who did not tacitly have these formal aims would not be engaged in volition, but some other kind of activity. Put another way, rational agents understand themselves to act from maxims that meet a condition of universal validity. If so, then the FUL is the principle through which one exercises the will and that tacitly guides all exercises of volition in its cognitive moment.

In short, for the will to be a law to itself, it must be part of the self-consciousness of rational volition that it has a formal aim that generates the internal formal principle through which the capacity is exercised, a principle that both describes and regulatively governs volition. For this

[17] For discussion of different conceptions of the will, see Gary Watson, "The Work of the Will," in Watson 2004. What he calls 'externalist conceptions of agency' separate normative judgment and decision about action (volitional commitment) and make the latter the work of the will. Kant's conception of agency is clearly internalist in this sense.

[18] Commentators who have defended some such interpretation of Kant's conception of volition include Herman 2007; Engstrom 2009; and Korsgaard 2009. Hill 2002: ch. 8 has expressed skepticism about this reading.

law to be the moral law (FUL), volition must be constitutively aimed at good. I should add further that the will is a law to itself in an interesting sense only if its formal principle generates substantive ends and principles that can initiate practical reasoning and determine the will to action.[19]

A second general point concerns the question, "what has autonomy?". If volition has two analytically separable moments, the cognitive/rational and the causal, where are we to locate its autonomy? A common answer is that autonomy is a property of the faculty of volition as a whole, specifically that *Wille* (the legislative function) is a law to *Willkür* (the executive function). (Cf. Beck 1960: 196–97, 199; Allison 1990: 130–31; and Timmermann 2007: 115, 174–45.) I'm inclined to agree that autonomy is a property of the will as a whole because the distinction between its two moments is an abstraction. But it can be misleading to say that '*Wille* gives the law to *Willkür*.' First, this way of putting it treats the power of choice as the elective capacity to decide whether or not to follow practical reason and judgment, a capacity with no formal aim of its own. But if (as it seems to me) Engstrom's reading of this distinction is correct, this approach misconstrues the sense in which the power of choice is the causal or executive dimension of volition.

Furthermore, though Kant does not explicitly say this, the causal moment of volition has its own formal aim and is thus a law to itself in just the same way as the cognitive. In the causal moment of volition one understands oneself to be an agent with certain causal powers and to have the formal aim of realizing the objects of one's practical judgments. One realizes these objects through actions that take sufficient means and by adjusting one's ends to one's actual causal powers. In other words, it is by following the Hypothetical Imperative (or more generally the norms of instrumental rationality) that one exercises one's causal powers as agent, and this principle both describes and regulates the operation of this power. That makes the Hypothetical Imperative (the principles of instrumental rationality) the internal formal principle of practical self-consciousness as an agent, the law that the causal moment of volition is to itself.[20]

[19] Thus I assume throughout this essay that the so-called 'content problem' in Kant's ethics has been resolved – as I think is amply demonstrated by Onora O'Neill's path-breaking work on the Categorical Imperative. For a brief statement, see O'Neill 1989: ch. 5.

[20] Though Kant does not make this point explicitly, Korsgaard and Engstrom have made it on his behalf: see Korsgaard 2008: ch. 1 ("The Normativity of Instrumental Reason") and Korsgaard 2009: 68–80, where she argues that Categorical and Hypothetical Imperatives are the constitutive principles of volition, and Engstrom 2009: 33–44. But Kant's explanation of the analytic character of the Hypothetical Imperative does clearly indicate that he takes this formal principle to be part of one's practical self-consciousness as an agent: "in the volition of an object as my

If the Hypothetical Imperative is a formal principle that the will is to itself, it also applies unconditionally – to agents qua rational, and not in virtue of any particular interests or ends. What then gives the Categorical Imperative a different normative status? The special authority of the Categorical Imperative is due to its synthetic character – that it yields some substantive ends and principles of conduct that are independent of and take deliberative priority over an agent's contingent ends. The Hypothetical Imperative is analytic because it derives conclusions about action from commitments to an end, which includes the representation of oneself as the cause of the end, and therefore does not lead to any substantive prescriptions without information about an agent's given ends.[21] At a more general level, the Hypothetical Imperative is analytic of rational agency because of its conceptual connection to the causal dimension of agency, which must be a component of any conception of rational agency. Any conception that did not understand rational agency as a form of causality would not be a conception of rational *agency*. By contrast, it may be true that rational agency (in us, or in its most complete form) is constitutively aimed at good and that the Categorical Imperative is analytically derivable from this dimension of rational agency. But there is no clear inconsistency in a form of rational agency that does not understand itself to aim at good, in which all ends are given by desire rather than set through judgments about good (that take themselves to be universally valid). If so, it is synthetic a priori that rational volition (in us, or in its most complete form) does have this formal aim.

The faculty of volition as a whole, then, has autonomy because the different moments of volition each generate their own fundamental norm. That said, what is correct in the common thought that *Wille* gives the law to *Willkür* lies in the priority of the cognitive moment in volition (cf. Engstrom 2010: 49–50). Practical reasoning and judgment about ends initiates volition, and choice properly follows practical reason or will (in the narrow sense) since its charge is to realize the objects represented by

effect, my causality as acting cause, that is, the use of means, is already thought" (*GMS* 4:417). In willing, one represents oneself as an agent, i.e., as the cause of one's end, and one causes one's ends by taking some necessary means. So commitment to take some necessary means is analytically contained in the self-consciousness that guides one's agency, or to use Engstrom's terms, in the 'causal moment' of practical self-consciousness.

[21] This issue – why the Hypothetical Imperative is hypothetical – is resolved in Hill 1992: 26–32, whom I follow. Hill suggests that an imperative is hypothetical (a) if its normative force for an agent is conditional on the agent having a particular end or (b) if it yields no substantive prescriptions for an agent without information about the agent's ends. The Hypothetical Imperative, though it applies to any agent qua rational, is a hypothetical imperative because it satisfies the second disjunct.

the practical judgment as good. It is the cognitive moment that brings the causal moment of choice into play. Further, since the will is a law to itself in an interesting sense only if practical reason gives itself the Categorical Imperative, the latter remains the primary element of Kant's thesis of autonomy of the will.

A third point concerns the sense in which the will 'gives' itself a law. So far I have unpacked Kant's thesis that the will is a law to itself, or the source of its own fundamental principle, through the idea that the nature of rational volition (or practical reason) supplies its own internal or formal principle. The role played by self-consciousness in guiding volition suggests a further sense of 'giving law' that adds an element of activity to the picture. The basic idea is that subjects engaged in certain forms of rational activity understand themselves to have certain formal aims and are normatively guided by their self-consciousness of these formal aims. Through this self-consciousness, agents 'give themselves' the relevant formal principles that, because they are part of the subject's understanding of what they are doing, both guide or describe the activity and serve as its regulative norm. One might think that it is a general feature of rational activity that it is self-conscious in this way, and moreover that it is normatively guided by its own self-consciousness (of the nature of the activity and its formal principles). The spontaneity of rational activity is that it is normatively guided by its own self-consciousness.

To apply this model to volition, I've suggested (on Kant's behalf) that volition has the formal aim of practical judgments about good (with an implicit claim to universality) that can realize their own objects. In exercising the will, rational agents understand themselves to have this complex formal aim – since this is what it is to will – and an agent who did not would not be exercising the power of volition. This self-consciousness can be expressed through the formal principles of volition that define the capacity, and because they are rooted in the self-consciousness of rational volition, tacitly guide all exercises of volition. Further, they function as regulative norms that set authoritative standards of normative success (the standard of means–end rationality, of universal validity), again because they are based in an agent's practical self-consciousness. Their normative hold on an agent comes from what one understands oneself to be doing in exercising one's will. In a word, agents 'give themselves' the formal principles that specify the nature of volition through the practical self-consciousness that guides volition.

One final comment to close this section: the structure of rational volition as I have laid it out makes plain the ways in which it is self-determining.

Each moment of volition is the source of its own formal principle, and individual agents 'give themselves' these formal principles through their tacit understanding of its formal aims. Furthermore, the formal norms are internal principles that guide the activity of volition through its own self-consciousness, without the need for any external influence. So practical reasoning moves from rational principle to practical judgment (about ends) guided by self-consciousness of its own formal principle, without the need for any additional impulse from the outside. Practical reason may need material from sensibility about which to reason, but its reasoning (both to and from ends and substantive principles) is determined by its understanding of its own principles (independently of sensibility).[22] Likewise, the power of choice operates through its own formal norm, without the need of any sensible interest to move it along. From a representation of a given end, choice moves to a representation of an action within one's power that will achieve the end through the representation of oneself as a cause, and this representation guides the deployment of an agent's causal powers. As noted above, without a representation of some end as good, there is no representation of an object to actualize, nothing for the causal moment of volition to work with. But given such an end, the power of choice guides itself through understanding of its own formal norm.

This is simply to say that each aspect of volition, and thus volition as a whole, is a spontaneous and self-active capacity that is normatively guided by its own self-consciousness. We might even say that rational volition necessarily proceeds under the Idea of freedom. Its judgments (about the goodness of ends and actions, judgments that derive actions from principles) do not "consciously receive direction from any other quarter" – they are not determined by any outside influence, but only by application of the internal principles of rational volition. And it is "the author of its own principles independently of alien influence" (*GMS* 4:448) in that it is the source of its own formal principles, and their normative authority for individual agents comes out of the practical

[22] I don't wish to commit to the view that pure practical reason does need material from sensibility in order to arrive at substantive practical principles, only to allow that this is a possibility that would not undermine its 'purity.' For pure reason to be practical, it has to generate some ends and principles that can determine volition, such as the standard moral principles and ends of virtue. Here I simply wish to leave open the possibility that to arrive at such principles, pure practical reason may need some input from sensibility, such as the interest in happiness or in fulfilling certain basic needs. It is important to distinguish between sensible interests providing some content about which to reason and sensible interest driving the reasoning, and only the latter needs to be excluded.

self-consciousness that guides volition (cf. *GMS* 4:448). Rational volition is a free capacity.

4 BUILDING OUT

I began by proposing a normative reading of Kant's thesis that the will, or practical reason, is a law to itself as the sovereignty of the will over itself, and I have interpreted autonomy of the will through the idea that practical reason has a formal principle given by its nature and not by any specific object that one wills. When understood in this way, we can see why autonomy of the will is needed to explain the normative authority of practical reason. The authority of the principles of practical reason comes from their internal role in specifying the capacity for rational volition and from the practical self-consciousness that guides volition. These principles are the internal formal principles of rational volition, and agents 'give' themselves these principles through their practical self-consciousness, their understanding of what it is to exercise one's will. That is to say, roughly, that the authority of the principles of practical reason comes from one's self-conception as a rational agent, not from any specific object that one wills. But certainly this narrowly focused notion of autonomy is related to other ideas in Kant. In this section, I'll consider how to build out to the capacity to act on the principles of pure practical reason independently of sensibility (the motivational independence of the will), to transcendentally free agency, and to the sovereign status of individual agents. These other notions all presuppose and follow from the fundamental idea of autonomy. Motivational independence of the will is the capacity underlying the causal moment of rational volition. Freedom of the will can be understood either as this motivational capacity, or more broadly as a consequence of the autonomy of the faculty of rational volition as a whole. And the sovereign status of individual agents is a normative implication of the idea that the will is a law to itself.

The capacity to act from principles of pure practical reason is a motivational capacity, specifically to determine one's causal powers through representations based in pure practical reason (rational principles or practical judgments), independently of sensibly given interests. To act on such representations, practical reason must produce them. So this motivational capacity presupposes, first, that the rational will, or practical reason, is a law to itself (independently of the property of any objects of volition) as explained above. Second, this motivational capacity obviously requires the capacity to act on the judgments that issue from pure

practical reason – the capacity to move from these practical judgments to representations of specific actions as good that guide the deployment of one's causal powers, without the need of any further sensible interest. The capacity to act on the judgments of practical reason is secured by the fact that the will is a faculty of *desire*, or *practical* reason that operates through its own internal principles. Realizing the objects of its desiderative representations is its formal aim (given by its nature without appeal to any particular object of volition). And because it has its own internal principle of activity, it operates without the need for any external (sensible) motivational influence to move it along. The motivational capacity to act from pure practical reason independently of sensibility, then, presupposes that practical reason gives itself the Categorical Imperative and is just the capacity underlying the causal moment of a volitional power that gives itself and functions according to its own formal principle.

Kant appears to offer us two ways of connecting autonomy and freedom of the will. He defines freedom of the will as the capacity to act on the form of universal law (on the law that the will is to itself), independently of determination by sensible interests (*GMS* 4:446; *KpV* 5:28–29, 33; *RL* 6:213–14, 226). Coupled with the idea of the will as a rational causality whose freedom "would be the property of such causality that it can be efficient independently of alien causes determining it" (*GMS* 4:446), these definitions suggest that freedom of the will is the motivational capacity just discussed – the capacity underlying the causal moment of volition.[23] In this case, it is a capacity that presupposes but appears a bit farther downstream from the narrowly focused conception of autonomy.

However, there are passages where Kant identifies freedom of the will and autonomy: "What, then, can freedom of the will be other than its autonomy, the property of being a law to itself?" (*GMS* 4:447); "Thus the moral law expresses nothing other than the autonomy of pure practical reason, that is freedom" (*KpV* 5:33). Presumably in these passages Kant means to identify the autonomy and the freedom of a unified faculty of volition whose different moments have their own interrelated formal ends and principles. In this case, the freedom of the will (in the broad sense) is indeed its autonomy, but it is a consequence of the self-determining structure of a will that gives itself its own formal principles. Free will is a

[23] In the *Groundwork* and the second *Critique*, Kant talks about freedom of will – by which he presumably means *Wille* in the broad sense. In the *Metaphysics of Morals* and the *Religion*, he refers to freedom of the power of choice [*Willkür*], and in the former says that only *Willkür*, not *Wille*, can be called free (cf. *RL* 6:226). Identifying freedom of the will with the capacity underlying the causal moment of volition fits Kant's language in the later works.

self-determining causality, a capacity to effect changes and initiate series of events 'from itself' that is independent of empirically given causes.[24] A volitional capacity that is a law to itself in both of its moments and that functions according to its own formal principles, without the need for any external motivational influence, satisfies this definition of free causality. The second of these approaches is more complete and subsumes the first, since the will is not fully self-determining unless its formal principles generate some substantive ends and principles to serve as starting points of practical reasoning – unless the formal principle of practical reason supports some substantive conclusions about ends and actions. But either way, the idea of autonomy drives the idea of free agency and supplies its positive conception.

Kant's thesis of autonomy requires a foray into some of the more abstract regions of his thought about the normativity of reason that are somewhat distant from the concerns of ethics, but we should not lose sight of its practical implications. Kant takes the autonomy or sovereignty of the will to lead fairly directly to the sovereign status of individual agents as ends in themselves – a status describable as their moral autonomy. Although (curiously) the idea of rational nature as an end in itself plays no explicit role in the argument of the second *Critique*, this work contains a nice statement of this idea, which it presents as a consequence of autonomy of the will.[25] Without trying to develop a tight argument, here is one way to lay out this connection. Given that the will is a law to itself, agents with the power of rational volition are self-determining. They can determine their activity through practical judgments of good and have the capacity to make assessments of universal validity. Conduct that affects an agent with the capacity for self-determination "is restricted to the condition of agreement with the *autonomy* of the rational being, that is to say, such a being is not to be subjected to any purpose that is not possible in accordance with a law that could arise from the will of the affected subject himself; hence this subject is to be used never merely as a means but as at the same time an end" (*KpV* 5:87). That is, treatment of and attitudes towards

[24] Cf. *KrV* A446/B474, A533/B561, where Kant defines freedom in the transcendental sense as "an **absolute** causal **spontaneity** beginning **from itself**" and "the faculty of beginning a state from itself, the causality of which does not stand in turn under another cause determining it in time in accordance with the law of nature."

[25] At *KpV* 5:87. Late in *Groundwork* II (*GMS* 4:435–40) Kant likewise suggests that status as end in itself is a consequence of autonomy of the will. See in particular 4:435 – a rational being's "share" in giving universal law confers membership in a kingdom of ends – and 4:440 – the moral agent has dignity not as subject to the moral law but "only insofar as he is law-giving with respect to it and only for that reason subordinated to it."

self-determining agents are to be governed by principles that could arise from their own will, where the standard of what could arise from their will is the condition of universal validity expressed by FUL. Agents who are to be treated according to this principle are *ipso facto* ends in themselves, and this status follows from the autonomy of the will.

CHAPTER 3

Vindicating autonomy: Kant, Sartre, and O'Neill

Karl Ameriks

1 VINDICATING KANT AND O'NEILL IN A BROADER CONTEXT

Kant's notion of autonomy remains at the center of contemporary debates, both as an object of attack and as a rallying point for spirited defenses of strict conceptions of morality. Among recent works in defense of Kant, Onora O'Neill's writings have long stood out for so clearly expressing Kant's views in a way that is highly relevant to contemporary philosophy and yet also avoids many of the controversial departures that characterize *broadly* 'Kantian'[1] approaches. O'Neill frequently contrasts Kant's own notion of autonomy with common approaches that exaggerate either one of the two core features that arise from the very structure of the term. The "auto" characteristic concerns its self-directedness, that is, *independence* and freedom in a primarily negative sense; the "nomos" characteristic concerns its *lawfulness*, that is, rationality and freedom in a primarily positive sense.[2]

O'Neill characterizes two significant misunderstandings along these lines with the labels "radical existentialism" and "panicky metaphysics."[3] The first misunderstanding corresponds to the thought that autonomy has to do with making an absolute value of choice for its own sake, so that 'autonomous' acts are found wherever one acts to express oneself as such, with no special regard for any content, let alone law, that the act might serve. The second misunderstanding corresponds to readings of Kant as a rigoristic rationalist who invokes a wholly transcendent 'metaphysical self' that demands we act in a 'Prussian' fashion and serve any

[1] See, e.g., Onora O'Neill, "Kant's Justice and Kantian Justice," in O'Neill 2000: 65–80.
[2] O'Neill, "Agency and Autonomy," in her 2000: 29.
[3] O'Neill 2000: 39, 43; and cf. her "Vindicating Reason," in Guyer 1992: 299–300. I do not discuss her criticism of "empiricist" approaches to Kant, even in Rawls, but it indicates openness to rationalism in a broad sense.