

1 Introduction

Immanuel Kant claims that the fundamental principle of morality is given by pure reason itself. His *Groundwork of the Metaphysics of the Morals*, published in 1785, starts with the statement that “a law, if it is to hold morally, that is, as a ground of an obligation . . . must not be sought in the nature of the human being or in the circumstances of the world in which he is placed, but *a priori* simply in concepts of pure reason” (G 4: 389). He later grounds its central argument that the moral law does apply to us human beings on the claim that we do have reason: “a human being really finds in himself a capacity by which he distinguishes himself from all other things, even from himself insofar as he is affected by objects, and that is *reason*” (G 4: 452). The *Critique of Practical Reason*, published in 1788 in order to defend the *Groundwork*, states that its “first question” will be whether “pure reason of itself alone suffices to determine the will or whether it can be a determining ground of the will only as empirically conditioned,” and decisively opts for the first of these alternatives: “reason can at least suffice to determine the will and always has objective reality insofar as volition alone is at issue” (CPrR 5: 15). By this Kant means, first, that pure reason by itself suffices to furnish the fundamental principle of morality by which the human will *ought* to be determined – in Kant’s model of human action, in the agent’s selection of particular maxims, suggested by experience, upon which to act in particular circumstances – and, second, that pure reason, by itself, can suffice to *motivate* human beings to act as they ought to act.¹

¹ Kant’s division of the cognitive powers or “faculties” of human beings evolved over his career, and his use of his own terminology was often flexible. His most fundamental distinction was between sensibility and intellect, with the former being our receptivity to representations from external objects or internal states and the latter our ability to organize our thought about such representations. Our most immediate representations of particular objects are called intuitions (*Anschauungen*), and our general representations of objects by means of marks that particular objects may share are called concepts. From the *Critique of Pure Reason* (first edition 1781) onward, Kant generally divides intellect into understanding and reason: understanding is the ability to form concepts and apply them either to particular objects or to other concepts in judgments, although in the *Critique of the Power of Judgment* (1790) Kant elevates judgment to a separate faculty; reason is, in the first instance, the ability to concatenate judgments, for example, in inferences. Sensibility, understanding, judgment, and reason all have characteristic forms that are not derived from experience but are applied to experience through empirically given intuitions: the pure forms of space and time in the case of sensibility, the pure categories of the understanding, and the pure ideas of reason, namely first the forms of inference and then the unconditioned ideas of the soul, the world-whole, and God. We can have *a priori* knowledge of the structure of the pure forms of sensibility and understanding, and these forms are also the forms of empirical knowledge. If reason is applied to our representations of how things are in the form of intuitions, concepts, and judgments, it is theoretical reason; if it is applied to our representations of how our own actions could bring about what ought to be, it is practical reason. The attempt to know what is through the theoretical use of reason alone would be speculative reason, and in Kant’s view is a failure, because the unconditioned ideas of reason by their very nature outstrip anything that can be given in experience; however, the attempt to determine how we ought to act

Sometimes Kant seems to go even further than these claims, already bold enough, to assert that the human will and pure reason are *identical*, as when he says that “the will is nothing other than practical reason” (G 4: 412). This remark has led many to ask how Kant could possibly explain voluntary but *immoral* behavior if he identifies reason as the source of *both* the moral law and all willed action: How could a will that as pure reason gives itself the moral law then act on any ground other than that?² But Kant clarifies this statement so quickly that no one should be misled by it, for when he makes it he is explaining why principles of reason, valid for all rational beings, present themselves *to us human beings* as imperatives:³ “If reason infallibly determines the will,” he continues, then “the will is a capacity to choose **only that** which reason independently of inclination cognizes as practically necessary, that is, as good,” but “if the will is not **in itself** completely in conformity with reason (as is actually the case with human beings),” then the will is *not* identical with practical reason and what pure reason requires of us can appear to us as a constraint or “necessitation” (*Nötigung*) (G 4: 412). Human beings can have inclinations – empirically given desires – toward actions contrary to what morality requires, so what morality requires can come across to us as a constraint. Likewise, the passage quoted from the *Critique of Practical Reason* implies only that pure reason is *capable* of determining how the human will *should* act, hardly that it determines how the human will or human being always *does* act. To be sure, even this more limited claim was a bold one for Kant to make when David Hume had argued that moral principles must be grounded in sentiment rather than reason precisely because “morals” must “have an influence on the actions and affections” and “reason alone . . . can never have any such influence.”⁴ Later we will consider some aspects of Kant’s theory that pure reason is capable of moving creatures like us even though we

by reason alone would be the pure use of practical reason, or pure practical reason, and because it does not purport to tell us how the actual world is but how it ought to be, it is in Kant’s view a success. Finally, the human ability to determine action is called “will” (*Wille*), but Kant will ultimately divide that into two parts: the ability to furnish principles of action (*Wille*) and the ability to choose which principles to act upon (the power of choice or *Willkür*). Pure *Wille* is identical to pure practical reason and provides the moral law. For Kant’s distinctions, see especially CPR A19-20/B33-4, A298-302/B355-9, and A320/B376-7, as well as CPrR 5: 37 and 15–16.

² This question was raised in Kant’s own time by Johann August Heinrich Ulrich (1746–1813) in Ulrich 1788, and a hundred years later by Henry Sidgwick in Sidgwick 1888. For contemporary discussion, see Wuerth 2014, ch. 7, and Guyer 2018a.

³ The validity of the moral law for all rational beings does not commit Kant to the actual *existence* of any rational beings other than human beings; its validity for other *possible* rational beings is intended to guarantee that the moral law is not grounded on merely contingent aspects of human nature.

⁴ Hume 1739–40, Book 3, Part 1, Section 1, paragraph 6.

have all sorts of desires to act otherwise than as morality demands. But our first question is simply, how did Kant think that pure reason of itself can provide the fundamental principle of morality? Or, more fully, how did he think that reason can determine the fundamental principle and the proper “object” of morality, that is, the goal of morality? How can reason alone determine even what we *ought* to try to do and accomplish in the name of morality? What *is* reason, in Kant’s view, that it can do *that*?

Kant does talk about “practical reason.” Obviously: The title of his second critique is the *Critique of Practical Reason*, and its aim is “to show that there is pure practical reason, and for this purpose it criticizes reason’s entire *practical faculty*” (CPrR 5: 3). By practical reason Kant means reason insofar as it bears on volition and action, thus on our choice of principles for action and our motivation by those principles, and his argument that there *is* such a thing as pure practical reason is intended to show that the application of reason to action is not limited to providing merely technical advice⁵ on what means to use to achieve ends that are set for us by desire, as Hume had asserted in his *Treatise*.⁶ Kant’s position is that reason can give us moral principles and set our moral goals on its own, as well as motivate us to act in accordance with these. But Kant also insists that “there can, in the end, be only one and the same reason, which may be distinguished merely in its application” (G 4: 391). That is, Kant does not suppose that practical reason is a faculty distinct from theoretical reason, or that practical reason has a special form or special forms for reasoning about action that can be understood apart from our forms for reasoning in general. For Kant, the principles that determine how we should act are fundamental to reason as such.

To be sure, there are differences between the application of reason to matters of fact and to matters of action. For one, in the theoretical use of reason we reason about how things *are*, while in the practical use of reason we reason about how things *ought* to be: “insofar as there is to be reason . . . something must be cognized *a priori*, and this cognition can relate to its object in either of two ways, either merely **determining the object** and its concept (which must be given from elsewhere), or else also **making** the object actual. The former is **theoretical**, the latter **practical** cognition of reason” (CPR Bix-x). For another, and this is the conclusion of Kant’s entire philosophy, there are things the reality of which cannot be asserted on the basis of a strictly theoretical use of reason, namely the existence of God and of our own freedom and immortality.

⁵ Kant calls technical advice given by (empirical) reason on how to realize goals that are set not by reason itself but by desire “hypothetical imperatives,” “imperatives of skill,” or “technical” imperatives (G 4: 416–17).

⁶ See Hume 1739–40, Book 3, Part 1, Section 1, paragraph 12.

The theoretical use of reason is restricted by the limits of sensibility, that is, by what we can actually perceive, and we cannot perceive God or our own freedom or immortality; yet, Kant holds, we are nevertheless entitled to affirm the existence of these things on the basis of the practical use of reason. This is because he thinks that it “would be absolutely impossible” for us to act in accordance with the fundamental principle of morality if we could not “presuppose” the existence of freedom, God, and our own immortality as necessary conditions for the possibility of moral success, and these things are in any case not *disproven* by the theoretical use of reason (CPR Bxxviii); or, if “practical reason has of itself original *a priori* principles with which certain theoretical positions are inseparably connected . . . then it is clear that, even if from the first perspective [reason’s] capacity does not extend to establishing certain propositions affirmatively, although they do not contradict it, **as soon as these same propositions belong inseparably to the practical interest** of pure reason it must accept them.” In this sense the practical use of reason has “primacy” over its theoretical use, “assuming that this union is not **contingent** and discretionary but based *a priori* on reason itself and therefore **necessary**” (CPrR 5: 121). So there is certainly something distinctive about practical reason for Kant, namely, that once the fundamental principle of morality and its necessary goal have been determined, we are entitled to hold beliefs about the conditions for the successful realization of morally mandated action that we would not be entitled to hold on theoretical grounds alone. To that extent it makes sense to talk of practical reason as a distinctive form of reason in Kant’s theory.

But Kant claims this special entitlement for practical reason to affirm the conditions of the possibility of realizing the goals that morality sets for us only *after* he has derived the fundamental principle of morality from reason as such, not from any special kind of reason but simply from the application of the principles of reason in general to the case of action. One and the same reason that gives us the most fundamental principle of principles for thinking about what is, Kant claims, also gives us the most fundamental principle for deciding how we may and must act. Our first question, then, is how does Kant think that reason as such yields the fundamental principle of morality?

2 Reasons, Reasoning, and Reason as Such

I am stressing Kant’s ambition to derive the fundamental principle of morality from the principles of *reason in general* because many philosophers have recently attempted to derive morality from conceptions of *practical reason* as a distinct form of reason. I will consider several examples of the latter approach in this section before turning to my own interpretation of Kant in the following

one. For example, Christine Korsgaard has stated that the fundamental thing that “arises from our rational nature” is “our need to have reasons.” Here she has defined rationality on the basis of an antecedent conception of reasons.⁷ But she has said several things about what a reason is or what it is to have a reason. In *The Sources of Normativity* (1996), she presented as Kantian the requirement that “Each impulse as it offers itself to the will must pass a kind of test for normativity before we can adopt it as a reason for action,” and “the test is one of reflective endorsement.”⁸ So a reason is not a reason all by itself, like a tree falling in a forest whether anyone hears it or not; rather, an inclination – any sort of desire to act that might naturally happen to present itself to a human being⁹ – toward an action becomes a reason for action only once it has been endorsed as such by an agent capable of a certain kind of reflection.¹⁰ That just pushes the question of what a reason is back to the questions, what kind of agent and what kind of reflection? In *The Sources of Normativity*, Korsgaard argued that Kant’s own test for what could be a moral principle, simply that it have “the form of a law . . . *All that it has to be is a law,*” can be improved.¹¹ Thus she argued first that agents reflectively transform impulses into reasons from the standpoint of some “practical identity,” such as that of a being “a member of a certain religion, a member of an ethnic group, a member of a certain profession, someone’s lover or friend, and so on,” each of which “identities gives rise to reasons and obligations”: If one thinks of oneself as having such an identity, then “[y]our reasons express your identity, your nature; your obligations spring from what that identity forbids.”¹² But some if not all of these identities do not and cannot give rise to universally valid reasons and obligations: all too obviously not all human beings are adherents of the same religion or ethnic group – if they were, many of the most savage moments of human history would never have occurred – neither is everyone, nor can they be, members of the same profession, and so on. So these kinds of practical identities were only a first step in Korsgaard’s argument; she went on to argue that reflectively endorsing any of these kinds of *particular* practical identities

⁷ Korsgaard 2009, p. 24. ⁸ Korsgaard 1996b, p. 91.

⁹ In his 1798 *Anthropology from a Pragmatic Point of View* Kant defines an inclination (*Neigung*) as a “habitual sensible desire” (Anth, §73, 7: 251). But in his writings in moral philosophy he often speaks of inclination without any suggestion that it must be habitual (e.g., G 4: 397–8, 400), instead defining it simply as the representation or thought of pleasure or satisfaction or the opposite from the existence of some object or state of affairs (CPrR 5: 21–3). I follow Kant in using the term in this more general sense.

¹⁰ This is what Henry Allison had previously called the “Incorporation Thesis,” the thesis that the mere occurrence of an inclination does not determine the will, but rather only its “incorporation” into a maxim of a rational agent does, that is, its “being taken by the agent (at least implicitly) as [a] sufficient reason . . . for action”; Allison 1990, p. 126.

¹¹ Korsgaard 1996b, p. 98. ¹² Korsgaard 1996b, p. 101.

depends upon recognizing that no matter what particular practical identity you endorse you must also endorse “your identity simply as a *human being*, a reflective animal who needs reasons to act and to live.” Korsgaard further argued, “And so it is a reason you have only if you treat your humanity as a practical, normative, form of identity, that is, if you value yourself as a human being,” and then she took human identity as such in *anyone* as something that gives reasons for *everyone* to treat that general identity in *anyone* as a fundamental source of normativity, as giving everyone reason to act in certain ways and not in others. “If this is right,” she concluded, “our identity as moral beings – as people who value themselves as human beings – stands behind our more particular practical identities. It is because we are human that we must act in the light of practical conceptions of our identity, and this means that their importance is partly derived from the importance of being human.”¹³

This argument explicitly makes one strong assumption about rationality, namely that what it is to be human is (at least in part) to require sound reasons for one’s actions; and it makes another strong assumption implicitly, namely that a genuine reason for *anyone* is a genuine reason for *everyone* – that my identity as a requirer of reasons in general, in which the normative force of any particular practical identity that I may recognize is grounded, is also a reason for anyone to recognize my right to act on my own reasons, whether they share or endorse my particular practical identity or not. Korsgaard is more explicit about this second requirement in a more recent book, *Self-Constitution* (2009). Here she argues that a creature who acted without reasons would not be a unified agent, a person, at all, only a “heap” of impulses, or, more realistically, that “it seems rather *obvious* that a formal principle for balancing our various ends and reasons must be a principle for unifying our agency . . . so that we are not always tripping over ourselves when we pursue our various projects, so that our agency is not incoherent.”¹⁴ To prevent incoherence among our projects, or the impulses that suggest them, reasons cannot be “completely particular,” as they would be if it were “possible to have a reason that applies only to the case before you, and has no implications for any other case.”¹⁵ But what Korsgaard infers from this is that a genuine reason cannot be merely *general* – that is, perhaps

¹³ Korsgaard 1996b, p. 121.

¹⁴ Korsgaard 2009, p. 58. In fact, it is far from *obvious* that a human being must possess a unified representation or conception of herself. Adrian Piper has argued that Kant’s view that a moral *agent* must be unified derives from his complex argument that a human *subject* must possess what Kant calls “transcendental unity of apperception,” a (second-order) representation (or the possibility of one; see CPR B132) that all her (first-order) representations constitute representations of a single, unified self, and that this requires the use of concepts; see Piper 2013, volume II. For my own interpretation of Kant’s concept of apperception and the transcendental deduction, see Guyer 1987, Part II.

¹⁵ Korsgaard 2009, pp. 72–3.

valid only for one or some agents, and perhaps only for some period of time – but must be genuinely *universal* – that is, valid for *any* person (at least in a certain kind of situation) *all* of the time. Reasons “are universal,” although “universality is quite compatible with – indeed it requires – a high degree of specificity”;¹⁶ that is, a reason need not, indeed could not possibly be valid for everyone in any kind of situation, but it must be valid for anyone in a certain kind of situation. Thus Korsgaard concludes that reasons are “considerations that have normative force for *me* as well as you,” and *vice versa*, and therefore reasons are by definition public reasons¹⁷ – that is how she gets the moral law requiring universal validity out of the conditions of rational *self*-constitution. This argument clearly depends upon the assumption that anything that is a reason is universally valid: Korsgaard does not use the notion of a reason as an undefined primitive, but does take the requirement that a reason is a universally rather than merely generally valid ground for, or consideration in behalf of action to be self-evident, and derives the moral law as the condition of coherent agency at all by means of this assumption. So she does define reason in terms of a certain conception of rationality, namely that rationality requires universal validity.¹⁸

Allen Wood forthrightly identifies a reason with a universally valid norm and defines reason in terms of such reasons. Thus he defines reason as “the capacity to think and act according to *norms*” and “[a] *reason*, in the widest sense of the term,” in turn, as “whatever counts as normative for beings with the capacity to give themselves norms and follow the valid norms they recognize.” Thus “Reason is the faculty through which we recognize beliefs, desires, or choices as grounded on something with normative authority”; reason or rationality is therefore simply the capacity to respond to reasons.¹⁹ But a norm is itself defined by the requirement of universal validity. For Wood defines reasons themselves – “as distinct from impulses or inclinations” – as “inherently objective or universal in their validity.”²⁰ Thus reason is defined by its demand for universal validity, indeed as the capacity to be determined by the requirement of universal validity itself, and reasons are then defined as considerations in behalf of action that satisfy the standard of universalizability, that is, being determinative for anyone in the relevant circumstances. Or to put it more generously, on Wood’s approach the concept of reasons turns out to depend upon an antecedent concept of reason after all, but reason is simply defined as our demand for universal validity. As Wood has put it more recently, “rational principles are always valid, valid equally for all rational beings. Their ultimate

¹⁶ Korsgaard 2009, p. 73. ¹⁷ Korsgaard 2009, p. 192.

¹⁸ For criticism of Korsgaard’s move from personal to universal validity, see Wuerth 2014, pp. 291–2.

¹⁹ Wood 2008, pp. 16–17. ²⁰ Wood 2008, p. 16.

validity is not dependent on anything (such as contingent desires or the choice of ends) that might distinguish one rational being from another.”²¹

That reason demands universal validity also turns out to be the key to Onora O’Neill’s approach, although she takes the notion of *reasoning* rather than of *reasons* as her starting point. O’Neill writes:

Two features frame all of Kant’s discussions of reason. The first is his insistence that there is no independently given “canon of reason” that sets the standard for human reason. The second is his thought that since we have not been given standards for reasoning we must construct them, and that this is a shared task, to be undertaken by a plurality of free agents.²²

Contrary to Korsgaard, and in my view *correctly*, O’Neill does not take this to be a characterization specifically of *practical* reason, but of reason in general; as she says, “reason’s principles” or “precepts must apply both to thinking and doing. Kant often stresses the basic unity of theoretical and practical uses of reason.”²³ But she then goes on to assert a thesis that is un-Kantian and in my view incorrect, namely that reason in general does not assume or need “antecedently established, ‘eternal’ standards,” but rather that we “invent or construct standards for reasoned thinking and acting, standards that have the sort of generally recognized authority that we would look to find in anything that could count as a requirement of reason.” It is “only when free agents discipline their thinking and acting in ways that others can follow [that] their thought and practice exemplify the fundamental, if meagre, requirements of reason.”²⁴ This purports to be a purely procedural conception of reason or rationality: Whatever beliefs or principles of action survive the thoroughgoing attempt to accept only beliefs or principles that others can “follow” or accept count as rational, and reason is nothing other than the activity of employing this process or the capacity to do so. “Self-legislation,” in turn, “is not then a mysterious phrase for describing the merely arbitrary ways in which a free individual might or might not think, but a characteristic of thinking that free individuals achieve by imposing the discipline of lawlikeness, so making their thought or their proposals for action followable by or accessible to others.”²⁵ Consequently “the only thought or action that can count as reasoned is that which we structure by imposing the ‘form of law’ – of universality,”²⁶ and what morality requires is simply that we impose this requirement of rationality on our proposals for action or, in Kant’s terms, our proposed maxims for action. O’Neill insists that “principles of reason and of logic are distinct,” although “logic is abstracted either from the use of the understanding or from that of reason,” so “its

²¹ Wood 2014, p. 43. ²² O’Neill 2004, p. 187. ²³ O’Neill 1992, p. 21.

²⁴ O’Neill 2004, pp. 187. ²⁵ O’Neill 2004, p. 189. ²⁶ O’Neill 2004, p. 189.

vindication would have to be derived from theirs, rather than conversely,²⁷ as if the vindication of reason itself could proceed without any antecedent principles.

But O’Neill’s characterization of reason as requiring that we make our beliefs and proposals for action followable by others belies that idea, since it treats the requirement of followability – universality – as a given – and if not given by reason itself, then by what? In other words, O’Neill treats not the concept of reasoning but that of reason (although not just practical reason) as the primitive, and explicates it in terms of the requirement for universality or universal validity. This is revealed when O’Neill makes “three substantive points” about reason: that “the discipline of reason is *negative*; second, it is *self-discipline*; third, it is a *law-giving*,” which entails that the “discipline of reason . . . is at least lawlike.”²⁸ The self-disciplining function of reason depends upon its demand for lawlikeness, because we humans are not always naturally disposed to satisfy that requirement. But this function of reason presupposes that it does require lawlikeness and can be defined as such.

O’Neill might seem to lend credibility to her claim that Kant’s conception of reason does not presuppose logic or any other “eternal” standard when she remarks that “he constantly rejects conceptions of reason, such as the Principle of Sufficient Reason, which supposedly give sufficient instructions for all thinking and acting . . . His insistence that ‘reason is no dictator’ reiterates the thought that there is no algorithm that fully determines the content of reasoned thought and action.”²⁹ Kant certainly does not suppose that the requirement of universal validity and the principle of sufficient reason are *sufficient* conditions for determining the full range of either human theoretical beliefs or of human duties; for the latter as well as the former we need further, empirical information about human nature and the human condition (see *MM*, Introduction, 6: 217). Moreover, for certain of our duties, those that he calls “imperfect” duties such as the duties to cultivate our own talents or assist others in their pursuits of happiness, Kant is clear that no rule can ever mechanically determine precisely how we should fulfill these very general obligations. But this does not mean that Kant does not treat formal principles of reason – very much including the principle of sufficient reason, as we will see – as *necessary* conditions of *reasoning* because they are the fundamental principles of *reason*. That is exactly what he does.

Korsgaard, Wood, and O’Neill thus all accept the requirement of universality as a fundamental standard of *reason* that constrains what can count as *reasons* or *reasoning* for us, in spite of having tried to define the former in terms of the

²⁷ O’Neill 1992, pp. 14–15. ²⁸ O’Neill 1992, pp. 27–8. ²⁹ O’Neill 1992, p. 28.

latter. There can be no doubt that Kant too thinks of the requirement of universal validity as intimately connected with reason as such – but since Kant thinks that the pure forms of sensibility (space and time) and the categories of the understanding (substance, causality, and so on) also give rise to universality, more precisely to necessity and universality, he cannot himself take the demand for universality to suffice to define reason because it is not unique to reason. Further, the requirement of universality must be applied to something specific in order to yield specifically moral results. This is recognized in the earliest of recent attempts to develop a moral philosophy inspired by Kant, that offered by Thomas Nagel in *The Possibility of Altruism* (1970). Nagel’s thesis is that “[a]ltruism . . . depends on a recognition of the reality of other persons, and on the equivalent capacity to regard oneself as merely one individual among many.”³⁰ His basic idea is that if one regards oneself from an “impersonal standpoint,” just as one person among all others, and regards any person as having good reasons for (some) actions, then one will recognize that one has just as much reason to promote anyone’s actions as one has to promote one’s own. Altruism will be the immediate consequence of the application of the demand for universality to the fact of being a person. But this leaves the concept of a person underspecified. Nagel recognizes that for Kant “[i]t is the conception of ourselves as free which [is] to be the source of our acceptance of the imperatives of morality,” and compares but contrasts this to his own approach dependent only on the concept of a person: “On Kant’s view . . . the agent’s metaphysical conception of himself” that occupies “the central role in the operation of moral motives . . . is that of freedom, but on [Nagel’s] view it is the conception of oneself as merely a person among others equally real.”³¹ Nagel clearly thinks that his foundation is less controversial and therefore more secure than Kant’s. But Kant’s view, I will suggest, is that the concept of a person has to be specified before it can play a foundational role for morality, and that it is to be specified precisely as that of an agent capable of setting his or her own ends. Applying the fundamental form of reason to the fact that persons are capable of setting their own ends is what will yield the fundamental principle of morality.

But it will also be central to my interpretation of Kant not that it is wrong to recognize that reason demands universality, as all of these versions of Kantian-style moral philosophy going back to Nagel have recognized, but that for Kant himself this demand is grounded in even more fundamental principles of reason, beginning with the law of *noncontradiction* – that is, the requirement to avoid self-contradiction as a condition of successfully asserting anything at all, let alone anything that others can follow or with which they can agree. Since

³⁰ Nagel 1970, p. 3. ³¹ Nagel 1970, pp. 12, 14.