

Kant: Rationality as Practical Reason

Onora O'Neill

Final Version published as *Kant: Rationality as Practical Reason* in The Oxford Handbook of Rationality, ed. Alfred J. Mele and Piers Rawling, Oxford University Press, Oxford, 2004, 93-109

Kant is famous for undertaking a critique of reason and for calling two of his most significant works *critiques of reason*.¹ These titles raise suspicions. Does Kant genuinely criticise reason, thereby calling into question the very processes by which any reasoned thought or action—including any criticism of reason—should be conducted? Or does he give these pretentious titles to works that deploy rather than criticise reason? Indeed, could anything really, seriously count either as a *critique* of reason or as a *vindication* of reason? Isn't the very idea that we could *show* that certain ways of thinking or acting are reasoned or reasonable absurd? After all, the demonstration must either build on assumptions that lack reasoned vindication, or be supported by arguments that deploy the very conception of reason supposedly vindicated. So it will be either unreasoned or circular: either way it will fail to vindicate reason. We have grounds for suspecting that no ways of organising thinking or acting have unconditional authority, and that Kant *cannot* have vindicated reason.

Kant's attempt to give an account of practical reason that offers unconditional reasons for action and provides the basis for a reasoned account of human duties, is spectacularly ambitious: even if it fails in some ways it is worth the closest attention. In this paper I aim to give as coherent account of that attempt as I can offer, although I shall say nothing about the connections Kant draws between practical and theoretical reason (See Neimann, 1994; O'Neill, 1989, ch 1 and 1992; Guyer 2000, ch 2;).

¹ Kant citations use the abbreviations indicated in the bibliography and the page numbers of the Prussian Academy of Sciences edition, which are given in the margins of each translation.

Since practical reasoning aims to shape and select action, I begin with a short account of Kant's views on action.

1. Practical Reasoning and the Agent Perspective.

Agents use practical reasoning to shape or guide their future action. Since practical reason has to bear on action yet to be done, it cannot bear on *act tokens*: there are no relevant, individuable act tokens at the time that practical reasoning takes place. So practical reasoning has to bear on *act types* (including *types of attitude*). It might be used to provide reasons for thinking that certain types of action or attitude are required or forbidden, recommended or inadvisable

As Kant sees it, types of action are specified by act descriptions, while normative claims are expressed in principles that incorporate act descriptions. Agents may consider, explore, test, adopt or reject practical principles. Kant speaks of the principle an agent adopts as *determining an agent's will*: it fixes—in the sense of *making determinate* rather than of *causing*—some aspects of the action or attitude to which an agent is committed. Kant calls the more significant determinations of an agent's will maxims (G 4:402, n., 4:421n; CPrR 5:19). Kant sees maxims as the practical analogues of *beliefs*. Individuals may *believe* a theoretical claim or proposition at or for some time; they may make a practical proposition their *maxim* at or for some time. Like beliefs, maxims have propositional structure and content, so are apt for reasoning. Kant's most basic thought about practical reason is that reasoning can bear on action because it is

formed or shaped by maxims, which have propositional structure and content.

In classifying only the more general principles that agents adopt as maxims, Kant is true to the etymology of the term. A maxim is the *maxima propositio*, a high or highest level proposition determining an agent's will at some time. The maxim an agent adopts will govern and inform other more specific decisions and aspects of his or her action or attitudes. For example, anybody who has adopted a maxim of not deceiving others is likely to express it in refraining from lying, in restraint in gossip, in care about checking facts and many other ways. Maxims can be for the long or the short term. They may be deeply entrenched in an agent's character or in the constitution of a collective agent (**R** 6:89), or adopted in face of a particular situation or for a short period. Kant usually speaks of agents as adopting a range of distinct maxims at a given time, but in a few passages he suggests that we can speak of the *deepest* or *most fundamental* principle of a person's character as a single maxim that governs their adoption of other more specific maxims, and thereby their entire life, often for a prolonged period.² However, maxims are always adopted and discardable, so something for which agents are responsible and which they might change.

There have been many discussions of the ways by which we may know Kantian maxims. (Herman, 1993; O'Neill, 1989 and 1996; Timmerman 2000). On some readings of Kant, agents are conscious of their maxims and know them by introspection. However this interpretation is hard to reconcile

² See the discussion of making either the moral law or self love one's most fundamental maxim and so determining one's basic disposition, **R** 6:22ff.

with Kant's views of the limits of human self-knowledge: he claims that we are not transparent to ourselves, but rather opaque. (G 4:406-12) Hence the notorious passages in which he points out that we cannot *know* whether there has ever been a *truly* loyal friend, or that we have acted *purely* for the sake of duty (G 4:407-8; Baron, 1995). On other views, maxims are ascribed to agents on the basis of a range of evidence, of which introspective evidence forms at most part. Both introspective and ascriptive views of knowledge of maxims focus on the *retrospective*, usually *third person* task of identifying the maxim(s) on which an act was done. In concentrating on the methods by which we can discover which maxim(s) are held by an agent at some time, they overlook the fact that the main task of practical reasoning is *prospective*. In reasoning about action we consider maxims that could be adopted, viewing them as principles or prescriptions for whose adoption reasons might— or might not— be found. A prospective and prescriptive account of maxims provides the basis not for discovering which maxim(s) an agent actually adopts on a given occasion, but for determining which maxim an agent has reason to adopt. The basic task of practical reasoning is to guide action rather than to adjudicate past acts.

Since Kant takes a prospective and practical approach to reasoning about action he can largely avoid the problem of showing how we are to discover agents' maxims, or work out the 'relevant' description for any act. When we try to assess action retrospectively, we have to work out which of many descriptions and principles satisfied by a given act is relevant for assessment. If assessment is undertaken for some specific purpose, such as financial audit or legal judgement, conformity or lack of conformity to relevant descriptions can be judged. And if the main aim of moral

assessment were to judge agents' maxims retrospectively, we would apparently need a general way of finding out what maxims they have 'really' adopted. Kant says a good deal about retrospective judgement of action in his discussions of reflective judgement,³ but (unlike some leading contemporary writers on moral perception, appraisal or judgement (Wiggins, 1987; McDowell, 1996)) he assigns priority to prospective, practical reasoning rather than to a retrospective, spectator perspective on action or ethics.

The potential maxims on which agents may bring practical reasoning to bear may be of many sorts. They do not have to be morally admirable: for example, Kant discusses a group of hard-bitten 'sophistical' maxims of political expediency.⁴ A neutral view of the moral status of maxims is appropriate: if practical reasoning is to show why we should adopt some rather other principles as maxims, setting prior limits on which principles are to be adopted as maxims would beg questions. Nor does Kant propose a method for ensuring that agents assemble possible maxims for consideration. Agents may well fail to consider some principles that they have reason to adopt as maxims. However, Kant takes it that agents are not likely to be systematically blind to the central principles of duty, which are repeatedly relevant to decisions and action.

³ In reflective judging 'only the particular is given, for which the universal is to be found'. **CJ** 5:180.

⁴ They include 'Fac et Excusa'; 'Si fecisti, nega' and 'Divide et impera'; **PP**, 8:374-5.

2 Hypothetical Imperatives

Some commentators have imagined that since Kant holds that practical reasoning can set unconditional requirements (so is a *Categorical Imperative*), he must deny that practical reasoning sets conditional requirements. Such a position would, of course, be absurd. Reasoning cannot guide action –cannot be *practical*—without taking a view of the connections between types of action and types of effect, between means and ends, between action and world.

Kant's account of instrumental reasoning is adjusted to his view that reasoning bears in the first instance on potential determinations of the will or maxims, and thence on action. He speaks of the fundamental principle of instrumental reasoning as the *Principle of Hypothetical Imperatives*, and formulates it as an abstract principle for rational willing. Hypothetical Imperatives “represent the practical necessity of a possible action as a means to achieving something else that one wills (or that is at least possible for one to will)” (G 4:414; also G 4:414-19; CPrR 5:19-20; Hill, 1992, chs. 1 and 7; Wood, 1999, ch 2). This principle requires commitment to the maxim “Whoever wills the end also wills (insofar as reason has decisive influence on this actions) the indispensable necessary means to it that are within his power” (G 4:417).

Kant argues that those hypothetical imperatives that supposedly guide the pursuit of happiness do not, strictly speaking, set *requirements* for action. Happiness is an indeterminate ideal—an “ideal of the imagination” (G 4:418)—so means-end reasoning directed at the pursuit of happiness yields at most approximate, *pragmatic imperatives* or *counsels of prudence*. These

commend ways of living that generally conduce to happiness, such as “frugality, courtesy, reserve, and so forth” (G 4:418). Other hypothetical imperatives are grounded in technical and causal requirements. They set genuine requirements, but only for those pursuing specific ends. Kant speaks of them as *technical imperatives* or *rules of skill*. Rules of skill are morally neutral: knowing the effects of poisons is as useful to physicians as it is to poisoners (G 4:415).

Kant's account of means-ends reasoning is clearly less ambitious than those favoured in many contemporary accounts of rational choice. He does not assume that we can list ‘options’ exhaustively, that we have complete or even very extensive knowledge of causal connections or probabilities, or that there is a metric for ranking or aggregating the value of ends. His account of instrumental reasoning does not provide enough structure for much in the way of judgements of efficiency, and says little about the distinction between willing necessary and sufficient means. He also does not place the entire burden of practical reasoning on means-ends reasoning.

3 Categorical Imperatives: Universal Law

In every area of life, instrumental reasoning is undertaken in pursuit of agents' chosen ends: these choices orient and require determinate means-ends reasoning. However, on Kant's account, specifically moral reasoning needs more than a combination of chosen ends and means-ends rationality. Those who hope to get by on this basis will defend some form

of ‘heteronomy in ethics’.⁵ They choose to make some intrinsically arbitrary ends the basis of their ethical reasoning. Their choices may variously endorse or defer to self-interest, religious dogma, established ideology, community ‘values’ or some version of self-development. Even those who propose an account of human flourishing or happiness as the foundation and context for moral reasoning do not escape this arbitrariness, given that they lack an adequately determinate account of human happiness or flourishing. Kant’s claims about the arbitrariness of heteronomous ethics are not hard to appreciate. Yet the thought that there could be any way of avoiding heteronomy in ethics looks quite implausible. How can there be *any* unconditional requirements on action? What reason have we to think that there is anything that could count as the rejection of heteronomy in ethics—as *autonomy in ethics*?

Kant’s answers to these questions can be found mainly in the Critique of Practical Reason and the Groundwork of the Metaphysics of Morals, although significant additional analyses and comments occur throughout his later writings on ethics, politics, religion and history. His thoughts are often difficult to follow, in part because he offers several seemingly distinct (yet supposedly equivalent) formulations of the supreme principle of practical reasoning, the Categorical Imperative. I shall begin with some comments on the well-known Formula of Universal Law [FUL], the formulation most closely linked to Kant’s attempted vindication of practical reason. Finally, I will comment on some of the other formulations and on Kant’s claims about their equivalence.

5 See the discussion on heteronomy and autonomy in ethics at **G** 4:440-44 and that of public and private uses of reason in **WE** 8:35-42.

FUL proposes a doubly modal requirement as the basis for moral reasoning. Its best-known version runs: *Act only in accordance with that maxim through which you can at the same time will that it become a universal law* (G 4:421). Many commentators—not least among them John Stuart Mill (Mill, J. S., 1962, 254)—have claimed that this is too little. Despite his protestations, Kant must in the end go beyond the modal demand set out in FUL if his conception of practical reason is to guide action. Surely almost any maxim that *can* be adopted by an individual agent *can* also be adopted by all agents? The only maxims that cannot be adopted by all are, it may seem, those that are intrinsically incoherent, so cannot really be adopted by individuals either (e.g. a maxim of being a popular recluse). Even maxims that refer to positional goods (e.g. a maxim of becoming richer than everyone else) could be *adopted* by all, although failure to achieve the maxim's aim would be guaranteed for all but the most successful agent. It seems that a requirement to act only on maxims that can be willed as universal laws cuts little ice.

Such criticisms oversimplify Kant's account of FUL. Kant views FUL as demanding that an agent be *able to will* the maxim he or she proposes to adopt 'as a universal law'. Willing is not merely a matter of *thinking* or of *wishing* some practical proposition: it is matter of making a certain proposition one's maxim, of adopting it as a determination of one's will, and this means engaging with the demands of the Principle of Hypothetical Imperatives (G 4:394). Willing a maxim 'as a universal law', although only a hypothetical test (I cannot *literally* will for others), requires agents to consider whether everyone *could make the proposed maxim the*

determination of his or her will, so must engage (hypothetically!) with the demands of the Principle of Hypothetical Imperatives.

When I will a maxim 'as a universal law' I do not (even hypothetically!) will that *everyone act on the maxim* at some moment, or at all moments. Very many practical principles, including principles of great and of trivial moral significance, cannot be acted on by everybody at any one time and place. If everyone tries to help a drowning child or to swim across a river simultaneously, overcrowding and mutual obstruction would guarantee that some cannot act. Kant's question is whether *everyone could will a maxim*. There is nothing incoherent in everyone *willing* to rescue a drowning child or to swim across a river, or even a particular river—but not everyone can act simultaneously on either principle in a given location. (Anyone who wills a child's rescue will remain on the shore if wading in would obstruct the rescue).

At this juncture it may seem that the demand that agents act only on maxims that they can will as universal laws fails to guide action for quite a different reason. Its problem is not that it is 'too specific', that it rules out each determinate act as impermissible on the grounds that nothing can be done by everyone at a given time and place. Its problem is that it is 'too general', since it requires only that agents can *adopt* maxims that could be adopted by all, so perhaps rules out nothing. Surely, if *anyone* can adopt a practical principle, then *everyone* can adopt it.

Kant thinks that this is not the case. Some principles can readily be adopted by a given agent or a minority of agents, *but only on the assumption that they are not adopted by all others*. For example, adopting a maxim of

promising falsely commits an agent to supporting means to promising falsely, hence to maintaining enough public trust for promises to gain acceptance. But willing false promising 'as a universal law' (*per impossibile*) commits an agent to willing the consequences of universal false promising, which include the destruction of trust, hence is incompatible with willing any reliable means to false promising—for oneself or for others. That is why Kant thinks that we *cannot* will false promising, coercion and many other types of action that victimise others 'as universal laws'. He points out that we do not even pretend to do this:

If we now attend to ourselves whenever we transgress a duty we find that we do not in fact will that our maxim should become a universal law—since this is impossible for us—but rather that its opposite should remain a law universally: we only take the liberty of making an exception to it for ourselves or even just for this once. (G 4:424)

But why is rejecting maxims that cannot be willed 'as universal laws' a formula for identifying *duties*? Even if we accept that FUL appears to show that some maxims of false promising cannot be 'willed as universal laws'—cannot be universalised—is this more than a curiosity? Why should we think that FUL picks out principles of action that we ought not to adopt? And even if it does, why should we think that it is a version of the 'supreme principle of morality'? And why, above all, should we think that this curious formula is the fundamental principle of practical reason? Further, why should we think that the other formulations of the Categorical Imperative, which Kant claims are equivalent to FUL, are also versions of the supreme principle of morality and of the fundamental principle of practical reason?

4 Universal Law as a Principle of Practical Reason

Before turning to the other formulations of the Categorical Imperative I shall consider why Kant sees FUL as (one version of) the supreme principle of practical reason. Kant's claims about practical reason can seem bombastic: phrases such as 'the supreme principle of practical reason' arouse suspicion. Yet Kant is very cautious about what reason, including the practical use of reason can provide.

This caution about the authority of reason is expressed in vivid terms in the prefaces and introduction of the Critique of Pure Reason. As Kant sees it, we do not even know where or how to begin the 'tasks of reason'. Reason is not *given* to us; it is not "whole and complete in each of us", as Descartes supposed (Descartes, 1985, Vol. I, 112). We constantly find ourselves using ways of thinking and acting that we speak of as reasoned; but we also find that daily reasoning goes horribly wrong. Particularly when we seek to extend our reasoning beyond experience, and aspire to reach metaphysical conclusions, we constantly find that

We have to retrace our path countless times, because ... [reason] does not lead where we want to go, and it is so far from reaching unanimity in the assertions of its adherents that it is rather a battlefield.... Still more, how little cause have we to place trust in our reason if in one of the most important parts of our desire for knowledge it does not merely forsake us but even entices us with delusions and in the end betrays us! (CPR Bxv; see also Aviii; Bxiv)

As is well known, Kant's central proposal for bringing the weary battles of metaphysics to an end is to insist that human reason cannot give us a route to knowledge that reaches beyond possible experience and the presuppositions of experience.

He insists that we must be wary of what we take to be the powers of human reason, and put them too to the test. Reason itself must be judged and scrutinised:

Reason should take on anew the most difficult of all its tasks, namely that of self-knowledge, and institute a court of justice by which reason may secure its rightful claims, while dismissing all its groundless pretensions... this court is none other than the **critique of pure reason** itself. (CPR Axii)

Yet the idea of vindicating reason by appeal to a 'court' or 'tribunal' can seem absurd. What procedures of judging could have the status to determine what does and what does not qualify and count as reasoning? Since there is no more general or fundamental claim to authority in organising thinking and acting than an appeal to reason, how can reason

itself be judged? How can any 'tribunal of reason' have standing to judge the competence and limits of reason? And how, if reason cannot be judged, can appeals to reason gain any authority? Perhaps the daring and demanding thought that reason lacks credentials leads to the conclusion with which postmodernists flirt: perhaps what passes for reason among us has no authority at all (O'Neill, 1989, chs 1,2).

Kant's move at this point⁶ has become more familiar during the last thirty years through the work of John Rawls, Jürgen Habermas, Thomas Scanlon and others. He proposes a *justification* or *vindication* of reason, rather than a *proof* or *foundation* for reason. Justification differs from proof in that it is directed to some audience: and unconditional justification must be directed to audiences without assuming that they meet any specific conditions, so must be directed to all agents. Procedures that can serve to reach a limited audience who share a particular conception of the world (perhaps embodied in shared beliefs or prejudices, in shared community or citizenship) can at best have restricted authority.⁷ At most they can provide a basis for parochial, conditional reasoning. By contrast, unconditional reasons must be fit to reach 'the world' (WE 8:38; in other translations 'the world at large'), rather than a restricted audience with whom an agent happens to share much. Reasoning that can reach only a restricted audience is incomplete or conditional: Kant calls it *private* or *heteronomous* reasoning.⁸ Reasoning

6 Kant makes the same move in his account of freedom, for which he offers a *vindication* or *defence* but no *proof*. See G 4:445-63.

7 Kant would therefore be unconvinced by John Rawls's conception of the reasonable as a form of public reason shared by fellow citizens within a bounded democratic society. He would think it inadequate as vindication of reason because it presupposes and does not justify bounded territories, citizenship and democracy.

8 See references in footnote 5.

that can reach the world at large is unconditional: Kant calls it *public* or *autonomous* reasoning. The fundamental move on which Kant's vindication of reason depends is the requirement that it be *fit for universal use*, rather than adapted to some restricted audience. By contrast, appeals to any local, restricted consensus or agreement would provide only parochial, limited and conditional reasons for action.

Those who aspire to offer reasons to unrestricted audiences, so assume no prior conditions that secure agreement, face a hard task. Their attempts at reasoning will fail unless they ensure that their proposals are accessible to an unrestricted audience.⁹ Those who propose *reasons to accept certain beliefs* to 'the world at large' must ensure that all others can in principle *follow* the moves that they make in presenting their thoughts: they must aim for intelligibility, without overtly or covertly assuming prior agreement. Those who propose *reasons for acting* to 'the world at large' must aim not only for intelligibility: they must propose principles of action that others not merely can follow in thought, but could adopt as principles of action. I do not offer reasons for action to all if I propose principles of action that I know some others *cannot* adopt. Another way of putting this requirement is to say that those who wish to offer practical reasons to an unrestricted audience must *Act only in accordance with that maxim through which [they] can at the same time will that it become a universal law* (G 4:421). FUL states (a version of) the supreme principle of practical reason because it states the condition for anything to count as a reason for action for an audience about whom we make no special, restrictive assumptions. It is

⁹ Accessibility is not the same as motivational sufficiency. For discussion of the motivational claims of Kant's view of practical reason see Korsgaard, 1996, ch 11.

therefore an requirement for giving unconditional reasons for action, a Categorical Imperative for the adoption of maxims. The underlying reason why Kant thinks that practical reasoning has to propose maxims that are fit to be universal laws is *that no other maxims can coherently be offered as reasons to all*. At most they could be coherently be offered as reasons to a restricted range of agents who accept *some further, rationally ungrounded assumption or attitude*.

5 Universal Law and Moral Duties.

If practical reason amounts to a demand to act only on principles that have the form of law, that can be principles for all, it offers only an *indirect* and *incomplete* standard for morality. I shall consider the implications of offering an *indirect* standard of morality in this section, and those of offering an *incomplete* standard in Section 6.

FUL provides only an *indirect* standard for guiding action since it identifies principles that ought to be rejected, rather than principles that ought to be adopted, or that it would be good to adopt. However, knowing that some principles ought to be rejected can guide action. If I know that a principle of revenge cannot be universalised, I have reason to reject a maxim of revenge. If I know that a principle of coercing others cannot be universalised, I have reason to reject a maxim of coercion.

Such reasons are, however, less than conclusive. It is sometimes impossible, hence not obligatory, to refrain wholly from types of action whose maxims

we have reason to reject. For example, Kant, as is well known, doubts whether human society can exist without some coercion: he is neither a pacifist nor an anarchist. Rather he argues that the very principle of rejecting coercion cannot be respected without using some coercion. A just political system not merely *may* but *must* coerce to limit coercion, and more generally hinder freedom in order to limit hindrances to freedom (MM 6:230-3; Mulholland, 1990; Flikschuh 2000; Timmons, 2002; Guyer, 2002). In acting on a maxim of rejecting coercion we have to deploy certain very specific forms of coercion. In identifying principles that we cannot universalise, so have reason to reject, Kant does not commit himself to principles of duty that are blind to circumstances and realities, or deny the possibility of conflicts between the various claims of (one or more) moral principles. (Herman 1993; Baron, 1995, ch 3; O'Neill, 2002a)

FUL can be used to identify further principles of action that cannot be willed as universal laws. Principles of doing violence, of victimising, or of undermining others' capacities to act in other ways, cannot coherently be willed as universal laws because their universal adoption (*per impossibile*) would predictably undercut the possibility of adopting those very principles for at least some others. (Herman, 1993; Baron, 1995; O'Neill 1989)

Those who adopt such principles in effect view themselves as enjoying exceptional moral status: they may know that their maxims cannot serve as principles for all, but see this as irrelevant because they do not view all others as their moral equals. On Kant's view, reasons for rejecting principles of action that cannot be universalised enable us to identify the fundamental principles of duty.

Kant divides basic duties into two classes, identified respectively by what he terms the *contradiction in conception* and *contradiction in the will* applications of FUL:

We must *be able to will* that a maxim of our action becomes a universal law: this is the canon of moral appraisal of action in general. Some actions are so constituted that their maxims cannot even be *thought* without contradiction as a universal law of nature... In the case of others that inner impossibility is indeed not to be found, but it is still impossible to *will* that their maxim be raised to the universality of a law of nature because such a will would contradict itself. (G 4:424)

The *contradiction in thought* (or *in conception*) test identifies maxims of *strict (narrow, perfect) duty*, including duties of justice. It picks out maxims of action that cannot coherently be thought of as principles for all. The contradiction in the will test identifies maxims of *wide (imperfect) duty*, including duties of virtue. It picks out maxims that can coherently be *thought of* as principles for all, but cannot be *willed as* principles for all in a world of interacting agents. For example, Kant thinks that *taken in isolation* a maxim of mutual indifference or a maxim of neglecting to develop any skills or talents could consistently be universalised: they pass the *contradiction in conception* test. But nobody can consistently will that these principles be universally adopted in any world of interacting agents whose members must (by the fact that they are instrumentally rational) will to receive others' help and to rely on others' skills if and when their own are insufficient: they fail the contradiction in the will test. (Herman, 1993, chs 3, 6 and 7; Baron, 1995)

The contradiction in the will test can be looked at in more than one way. On a minimal reading it is a matter of prudence. No reasonable agent who acknowledges *her own* finitude and vulnerability, and consequent inability to achieve all her ends unaided, can coherently will to be part of a world of agents who are indifferent to others' needs or who systematically neglect to develop human skills: in doing so she would flout the demands of instrumental rationality. On a wider reading, no reasonable agent who acknowledges *others'* finitude and vulnerability, so knows that nobody can achieve all their ends without help, can will universal indifference to human needs or to the development of human skills. Willing universal mutual indifference amounts to willing a world in which agency and capacities fail for some or many, so undermining action. For similar reasons, no rational agent can consistently will universal failure to develop skills. Willing universal failure to develop skills amounts to willing a world in which some or many find their capacities to act at risk. Kant, of course, acknowledges that some *individuals* may get away with large amounts of indifference to others, and with failure to develop skills: free riders often get away with it. He denies, plausibly enough, that we can will either sort of free riding as a universal law for a world of interacting agents.

6. Universal Principles and Judging Cases

Arguments from FUL to these broad principles of duty also offer a very *incomplete* standard for morality. The perennial allegation that Kantian practical reason is *too abstract* or *too formalistic* sees this incompleteness as a serious defect. (Mill, 1962, ch 1) The charge of *formalism* is that Kant

identifies only very general principles of duty, whereas we need to know just what to do in particular circumstances. Kant himself pointed this out:

A physician, a judge or a ruler may have at command many excellent pathological, legal or political rules, even to the degree that he may become a profound teacher of them, and yet, none the less, may easily stumble in their application. For, although admirable in understanding, he may be wanting in natural power of judgement. He may comprehend the universal in *abstracto* and yet not be able to distinguish whether a case *in concreto* comes under it. (CPR A134/B173)

Discussions of judgement, including practical judgement, are ubiquitous in Kant's writings. He never assumes agents can move from principles of duty, or from other principles of action, to selecting a highly specific act in particular circumstances without any process of judgement. He is as firm as any devotee of Aristotelian *phronesis* in maintaining that principles of action are not algorithms, and do not entail their own applications. There has been a good deal of recent discussion of the details of Kant's views on practical, including ethical, judgement. (Herman 1993; Engstrom and Whiting 1996; O'Neill 2002a) His discussions of these topics are numerous, complex and perhaps most abundant in The Metaphysics of Morals, which addresses many aspects of practical judgement, including deliberation, casuistry and conflicts of obligation.

A second version of the charge that Kant's ethics is seriously incomplete objects that it is possible to devise artfully tailored maxims that can be

willed as universal laws, but are morally obnoxious (Wood, 1999, ch 7; Herman 1993). For example, instead of testing a general maxim of revenge or false promising, as Kant does, we might test maxims permitting persons of specific sorts or status to exact revenge or to promise falsely, knowing that deception or revenge by narrowly specified categories of agents could be willed as universal laws without contradiction. Such proposals for undercutting the implications of Kant's ethics overlook two difficulties. First, they are based on according some people exceptional moral status, denied to others, so reject Kant's fundamental view that human beings are moral equals. Second, those who advance it fail to note that basic principles of duty do not fall away when we consider a more closely specified line of action. The general duty to reject revenge does not fall away just because a proposed act of revenge falls under more specific descriptions and principles; a general duty of fidelity does not evaporate because an agent is tempted by a scam that is open to few.

6 Universal Laws and Ends in Themselves

Perhaps the most mysterious feature of the Categorical Imperative is that Kant formulates it in several distinct ways that look quite different, but which he claims are equivalent (G 4:436; Hill 1991 and 1992; Korsgaard, 1996, chs 2-4; O'Neill 1989, ch 7; Wood, 1999, esp. ch 4). It is common to group the various formulations under four or five headings. Here I shall discuss the well-known *Formula of the End in Itself* (FEI) and *Formula of Autonomy* (FA), but say nothing specifically about the *Formula of the Law of Nature* (FLN) and the *Formula of the Kingdom of Ends* (FKE). There

are two reasons, considerations of space apart. Most immediately, I bracket FLN because it is in many respects similar to FUL, and FKE because it is readily understood if sense can be made of FEI. Secondly, and more importantly, FEI and FA are the origins of the resonant contemporary moral ideals of *respect for persons* and of *autonomy*. Yet it is far from clear that either FEI or FA is equivalent to FUL, or that either is a version of the supreme principle of morality, rather than one moral principle among others. Still less is it obvious why either should count as a version of the supreme principle of practical reason.

FEI is formulated in the Groundwork of the Metaphysic of Morals in the words *So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means* (G 4:429). Kant himself views FEI as the most intuitive version of the Categorical Imperative; many recent accounts of Kantian ethics go further and dismiss FUL in order to concentrate on FEI's demands for respect for persons (G 4:436; Hill 1992; Wood, 1999, ch 4; for criticism see Regan 2002). Certainly, FEI does not look equivalent to, or even like, FUL. Yet if no equivalent reading can be found, Kant's ethical theory will fall apart: the deepest arguments used to justify the Categorical Imperative are directed mostly at FUL, so can provide grounds for FEI only if it can be read as equivalent to FUL (O'Neill, 1989, ch 7).

The fundamental difference between FUL and FEI is that FUL constrains what agents should (may, may not) do, whereas FEI constrains how agents should (may, may not) be treated. The two formulae consider requirements on action respectively from the agent's and the recipient's point of view.

FEI, however, is more explicit than FUL about those on the receiving end of action. It requires right treatment *of humanity*,¹⁰ whereas FUL requires action on maxims that can be principles *for all*. If FUL and FEI are to be read as equivalent, it is essential to read them using a common account of the scope of ethical concern. For example, it is common to read FUL as requiring action on maxims that can be principles *for all human agents*, so using the same view of the scope of ethical concern as is explicit in FEI. (Alternatively, both FUL and FEI could be read as setting requirements on action by and treatment of rational natures, human or other. (Hill, 1992, ch 2).

Even when FUL and FEI are read using a common view of the scope of ethical concern, their equivalence is not obvious. Why should acting on principles that can be principles for all amount to treating others as ends and not as mere means? The issue can best be approached by considering first what is meant by 'treating others as mere means'. Kant distinguishes interaction with others that respects and preserves their capacities as agents, in which persons permissibly use one another as *a means*, from action that uses unwilling others as *a mere means*. We use others as mere means when we treat them not as agents, but as things or tools, as something to be "used by this or that will at its discretion" (G 4:428). Those who adopt such maxims do not always damage others' agency (they may lack opportunity or power) but the *standard* results of such action damage agency. We cannot will any of numerous maxims of victimising as universal laws: in doing so

10 Kant discusses the claims of *humanity*, of *rational nature*, and of *sentient nature*) in many passages, especially in G 4:448 onwards and in MM. For discussion of his humanism, and of charges of speciesism levelled against him, see Allen W. Wood and Onora O'Neill, 1998.

we would will results that would undercut the means for like action for some or for many. Those who treat others as *mere means* act on maxims that they *cannot* will as universal laws. This argument does not establish the equivalence of the FUL and FEI, but only that the first part of FEI—*do not use [others] merely as means*—follows from FUL. Strict duties can be derived either by applying the *contradiction in conception* test aspect of FUL, or by adhering to the part of FEI that prohibits treating others as mere means.

The second part of FEI enjoins treating others as ‘ends in themselves’, and corresponds to the contradiction in the will test application of FUL, by which maxims are tested not for inner coherence, but for fitness as laws for a possible world of agents. In treating others as ends in themselves, we treat them as persons, as beings who have objective worth: this requires more than refraining from treating them merely as means. We treat others as ends by acting in ways in which a world of agents can be sustained: by acting on maxims that can form part of a system of maxims that can be willed without contradiction, that harmonise with the necessary conditions for sustaining a world of agents (G 4:430-1).

Provided that FUL and FEI are read using a common view of the scope of ethical concern, the difference between them is in the end a matter of perspective. FUL is a test for *ways of acting* that could be adopted by all agents in a world of agents; FEI is a test for *ways of being treated* that could be accepted by all agents within some world of agents. Within a world of agents, the perspectives of agency and recipience consider the same set of possible actions from different points of view. Both formulae state

procedures for rejecting maxims whose universal adoption would undermine at least some others' possibilities of like action.

7 Universal Laws and Autonomy

In many ways the Formula of Autonomy (FA) is closer to FUL than it is to FEI. In the Groundwork Kant writes "*The principle of autonomy is... to choose only in such a way that the maxims of your choice are also included as a universal law in the same volition*"(G 4:439).¹¹ The puzzle is not so much that Kant sees FA as a version of the Categorical Imperative—it is, after all, quite close to FUL (but see Wood, 1999, Ch 5)—but rather to understand why he speaks of it as the formula or principle of *autonomy*. How can he view autonomy as basic to morality, or claim that "Morality is thus the relation of actions to the autonomy of the will"? (G 4:439)¹²

This puzzle can be resolved by distinguishing Kant's conception of autonomy adequately from contemporary conceptions. Most current discussions of autonomy see it as a property of individual agents, which they may have to a greater or lesser degree, and may express in some domains of life more than in others. Contemporary conceptions of autonomy generally equate it with forms of personal independence or self-expression, or with acting on certain distinctive, supposedly 'autonomous', preferences (O'Neill, 2002b, chs 2 and 4). Such conceptions of individual autonomy are widely discussed, and their ethical merits widely disputed. These

11 See Wood, 1999, 163-4 for a listing of versions of FA and an argument that FA is quite distant from FUL.

12 . Even more strikingly he equates it with practical reason: the power to judge autonomously-- that is, freely (according to principles of thought in general) -- is called reason" CF 7:27.

discussions are irrelevant to understanding Kantian autonomy. Kant never writes of autonomous *selves* or *persons* or *individuals*. He predicates autonomy *of reason, of ethics, of principles, of willing* (Hill, 1991, ch 4). Although contemporary advocates of individual autonomy often claim Kantian ancestry for their ideas, the claim is bogus, made plausible only by distorting Kant's conception of autonomy in major ways.

For Kant the idea of autonomy captures the two central aspects of his account of practical reason and of ethics: that duty is a matter of acting on principles or laws, and that those principles or laws should not be derived from arbitrary starting points. Principles are autonomous on Kant's view not because they express some particularly striking or independent personal decision or attitude, but because they are *not derived from elsewhere*. 'Self-legislation', as Kant writes of it, does not mean that each self or each agent chooses or 'legislates' moral principles for all (coherent only where there is some extraneous, hence unreasoned, source of coordination). He understands self-legislation as *the law giving of reason* rather than as *the law giving of individuals*. He remarks that

... it is requisite to reason's lawgiving that it should need to presuppose only *itself*, because a rule is objectively and universally valid only when it holds without the contingent, subjective conditions that distinguish one rational being from another" (CPrR 5:20)

and that states that autonomy requires a

“*law-giving of its own* on the part of pure and, as such, practical reason [which] is freedom in the *positive* sense” (CPrR 5:33)

For Kant the element *self* in the term *self-legislation* indicates a *reflexive* claim that lawgiving, and hence principles and maxims, not be derived ‘from elsewhere’, that is from sundry arbitrary assumptions or conditions.

Kantian autonomy is not a matter of self-expression by individual selves who ‘legislate’, but of agents choosing maxims—‘laws’—that are non-derivative, so non-conditional.

This conception of autonomy as willing that does not appeal to arbitrary (bogus, or at best conditional) ‘authorities’ is the basis of Kant’s contrast between *autonomous ethics* and positions that advocate forms of *heteronomous ethics*. The proponents of heteronomy in ethics derive supposed moral principles from ‘authorities’ such as Church or state, ideology or market forces, public opinion or personal preference, individual choice or majority vote.¹³ The practical reasoning deployed in heteronomous ethics is therefore essentially instrumental or conditional: it is simply a matter of choosing action that implements the standards of the supposed ‘authority’ effectively and efficiently.

As Kant sees it

If the will seeks the law that is to determine it anywhere else than in the fitness of its maxims for its own giving of universal law—consequently if, in going beyond itself, it seeks this law in a property of any of its objects—heteronomy always results. (G 4:441)

If we are to offer reasons for action that are relevant to *all* others, we must begin by ensuring that we offer reasons they *can* offer and consider, accept or refuse. We cannot do this by relativising our reasons for action to any 'authority' that some—or many—may have no reason to accept, and may in fact reject. Rather than trying to build personal independence or other conceptions of individual autonomy into ethics, Kant argues that we should ensure that what we propose to others is based on principles that are *fit* to be laws, hence on principles that they at least *could* adopt. This seemingly slender modal requirement demands that we reject all forms of heteronomous ethics in favour of acting on principles that are fit to be universal laws, hence also that we not treat others as mere means or as less than ends. Kant's sums up these views on reason and morality in the striking claim that "the moral law expresses nothing other than the *autonomy* of pure practical reason"(CPrR Theorem IV, 5:33):

Bibliography

- Allison, H., 1990, Kant's Theory of Freedom, Cambridge University Press, Cambridge.
- Baron, Marcia W., 1995, Kantian Ethics Almost without Apology, Cornell University Press, Ithaca, N.Y.
- Descartes, René, 1637, Discourse on the method of rightly Conducting One's Reason and Seeking the Truth in the Sciences, Philosophical Writings of Descartes, Vol. 1, tr. John Cottingham, Robert Stoothof and Dugald Murdoch, 1985, Cambridge University Press, Cambridge.
- Engstrom, S. and Whiting, J., eds., 1996, Aristotle, Kant and the Stoics, Cambridge University Press, Cambridge.
- Flikschuh, Katrin, 2000, Kant and Modern Political Philosophy, Cambridge, Cambridge University Press, Cambridge.
- Guyer, Paul, 2000, Kant on Freedom, Law, and Happiness, Cambridge University Press, Cambridge.

- Guyer, Paul, 2002, 'Kant's Deductions of the Principles of Right', in Mark Timmons, ed., Kant's Metaphysics of Morals: Interpretative Essays, Oxford University Press, Oxford, 23-64.
- Herman, Barbara, 1993, The Practice of Moral Judgement, Harvard University Press, Cambridge, Mass..
- Hill, Thomas E. Jr., 1991, Autonomy and Self-Respect, Cambridge University Press, Cambridge.
- Hill, Thomas E. Jr., 1992, Dignity and Practical Reason, Cornell University Press, Ithaca, NY.
- Kant, Immanuel, 1781, Critique of Pure Reason [CPR], tr. and ed. Paul Guyer and Allen W. Wood, 1996, Cambridge University Press, Cambridge.
- Kant, Immanuel, 1784, An Answer to the Question: What is Enlightenment? [WE], tr. Mary J. Gregor, in Kant's Practical Philosophy, 1996, Cambridge University Press, Cambridge.
- Kant, Immanuel, 1785, Groundwork of the Metaphysics of Morals [G], tr. Mary J. Gregor, in Kant's Practical Philosophy, Cambridge University Press, Cambridge 1996.
- Kant, Immanuel, 1788, Critique of Practical Reason, 1787, [CPrR], tr. Mary J. Gregor, in Kant's Practical Philosophy, tr. Mary Gregor, Cambridge University Press, Cambridge 1996.
- Kant, Immanuel, 1790, Critique of the Power of Judgement [CJ], TR. Paul Guyer and Eric Matthews, Cambridge University Press, 2000.
- Kant, Immanuel, 1793, Religion within the Boundaries of Mere Reason [R], tr. George di Giovanni, in Immanuel Kant, Religion and Rational Theology, ed. Allen W. Wood and George di Giovanni, Cambridge University Press, Cambridge, 1996.
- Kant, Immanuel, 1795, Toward Perpetual Peace, [PP], tr. Mary J. Gregor, in Kant's Practical Philosophy, Cambridge University Press, Cambridge 1996.
- Kant, Immanuel, 1797, The Metaphysics of Morals, [MM], tr. Mary J. Gregor, in Kant's Practical Philosophy, Cambridge University Press, Cambridge 1996.
- Kant, Immanuel, 1798, The Conflict of the Faculties, [CF], tr. Mary J. Gregor and Robert Anchor, in Immanuel Kant, Religion and Rational Theology, ed. Allen W. Wood and George di Giovanni, Cambridge University Press, Cambridge, 1996.
- Korsgaard, Christine, 1996, Creating the Kingdom of Ends, Cambridge University Press, Cambridge.
- McDowell, John, 1996, 'Deliberation and Moral Development', in Stephen Engstrom and J. Whiting eds., Aristotle, Kant and the Stoics, Cambridge, Cambridge University Press, 19-35.
- Mill, John Stuart, Utilitarianism, in M. Warnock, ed., Utilitarianism and other works, Fontana, 1962,
- Mulholland, Leslie, 1990, Kant's System of Rights, Columbia University Press, New York.
- Neiman, Susan, 1994, The Unity of Reason, Oxford University Press, Oxford.
- O'Neill, Onora, 1989, Constructions of Reason: Explorations of Kant's Practical Philosophy, Cambridge University Press, and Cambridge.
- O'Neill, Onora, 1992, 'Vindicating Reason', in Paul Guyer, ed., A Cambridge Companion to Kant, Cambridge University Press, Cambridge, 280-308.

- O'Neill, Onora, 1996, 'Kant's Virtues', in Roger Crisp, ed., How Should One Live? Essays on the Virtues, Oxford University Press, Oxford, 77-97.
- O'Neill, Onora, 2002a, 'Instituting Principles: Between Duty and Action' in Mark Timmons, ed., Kant's Metaphysics of Morals: Interpretative Essays, Oxford University Press, Oxford, 331-47.
- O'Neill, Onora, 2002b, Autonomy and Trust in Bioethics, Cambridge University Press, Cambridge.
- Regan, Donald E., 2002, 'The Value of Rational Nature', Ethics, 112, 267-291.
- Timmerman, Jens, 2000, 'Kant's Puzzling Ethic of Maxims', Harvard Philosophy Review, 39-52.
- Timmons, Mark, ed., 2002, Kant's Metaphysics of Morals: Interpretative Essays, Oxford University Press, Oxford.
- Wiggins, David, 1987, 'Deliberation and Practical Reason' in his Needs Values and Truth: Essays in the Philosophy of Values, Blackwell, Oxford, 215-37.
- Wood, Allen W. and O'Neill, Onora, 1998, symposium on 'Kant on Duties Regarding Nonrational Nature', Proceedings of the Aristotelian Society, Supp Vol, LXXII, 189-228.
- Wood, Allen W., 1999, Kant's Ethical Thought, Cambridge University Press, Cambridge.