KANT'S GROUNDWORK OF THE METAPHYSICS OF MORALS

A Commentary

Jens Timmermann

CAMBRIDGE

This page intentionally left blank

Kant's *Groundwork of the Metaphysics of Morals*

The *Groundwork of the Metaphysics of Morals* is Kant's central contribution to moral philosophy, and has inspired controversy ever since it was first published in 1785. Kant champions the insights of 'common human understanding' against what he sees as the dangerous perversions of ethical theory. Morality is revealed to be a matter of human autonomy: Kant locates the source of the 'categorical imperative' within each and every human will. However, he also portrays everyday morality in a way that many readers find difficult to accept.

The *Groundwork* is a short book, but its argument is dense, intricate and at times treacherous. This commentary explains Kant's arguments paragraph by paragraph, and also contains an introduction, a synopsis of the argument, six short interpretative essays on key topics of the *Groundwork*, and a glossary of key terms. It will be an indispensable resource for anyone wishing to study Kant's ethical theory in detail.

JENS TIMMERMANN is a lecturer in Moral Philosophy at the University of St Andrews.

Kant's *Groundwork of the Metaphysics of Morals*

A Commentary

JENS TIMMERMANN
University of St Andrews

CAMBRIDGE
UNIVERSITY PRESS

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

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/9780521862820

First published in print format 2007

ISBN-13 978-0-511-28907-1 eBook (EBL)
ISBN-10 0-511-28907-3 eBook (EBL)

ISBN-13 978-0-521-86282-0 hardback
ISBN-10 0-521-86282-5 hardback

For Bettina, Hans, Jakob, Ricarda,
Carlotta and Florentin

Contents

Commentary

Acknowledgements

I should like to thank the institutions that have made this book possible: the University of St Andrews for granting me sabbatical leave in 2005; the Arts and Humanities Research Council for awarding me back-to-back research leave in the second half of that year; the Royal Society of Edinburgh for funding a three-month research fellowship at the University of Göttingen, and the members the Göttingen department for making my stay so profitable and enjoyable. I owe thanks to Claudio La Rocca and Stefano Bacin for their wonderful philosophical hospitality at the university and at the Scuola Normale in Pisa during the summer of 2005, to audiences at Cambridge, Göttingen, Helsinki, Notre Dame, UC Riverside, UC San Diego and St Andrews, and to Onora O'Neill, Norbert Anwander, Fabian Freyenhagen, Andrews Reath, Oliver Sensen and Werner Stark for many enjoyable discussions of Kant's ethical theory. Particular thanks are due to my doctoral students at St Andrews for reading the complete manuscript: Caroline Benson and Ralf Bader, who also helped me to compile the index.

Note on quotations from Kant's works

Quotations from Kant's works have usually been adapted from the Cambridge Edition, published by Cambridge University Press under the general editorship of Paul Guyer and Allen W. Wood. The series includes Mary Gregor's translation of the *Groundwork*, and I generally follow the wording of that translation unless there is a good textual or philosophical reason to depart from it. I have also consulted the remaining seven or so English versions currently in use[1] and, on occasion, translations into other modern European languages.[2]

References are to volume, page and frequently line numbers of the standard German edition of Kant's works known as the Academy edition. Its twenty-nine volumes are published under the auspices of the Berlin-Brandenburg (formerly Royal Prussian) and Göttingen Academies by Walter de Gruyter in Berlin and New York. Academy page numbers are now commonly reprinted in the margins of other editions and translations. The reference 'IV 393' thus points to page 393 of volume IV of the Academy edition, 'V 97.19' to line 19 of page 97 of volume V. Readers who do not use this edition will still get a good impression as to where on a large, mostly thirty-seven-line Academy page the reference is to be found. As is customary, an exception is made for the *Critique of Pure Reason*, for which page numbers of the first (A) and second (B) editions are given: a reference of the format 'A 15/B 29' – simply 'A 361' or 'B 131' for material contained in the first or the second edition only – refers to the first *Critique*. Kant's handwritten notes or 'Reflections', consecutively numbered and printed in volumes XIV–XIX of the Academy edition, are quoted as 'R'. The lectures, contained in volumes XXIV–XXIX are quoted with reference to the name

1 In chronological order: Abbott, Paton, Beck, Ellington, Zweig, Wood and Denis's recent revision of Abbott's translation; see Bibliography.

2 Italian and French, but also curiosities such as A. Pannenberg's German (!) translation (Velhagen & Klasing, 1927).

of the student whose notes we possess: *Collins*, *Mrongovius*, *Vigilantius*, etc.

In the running commentary, the paragraph sign (¶), followed by volume, page and line number, indicates the beginning of a new paragraph in the text of the *Groundwork*.

Introduction

That light we see is burning in my hall.
How far that little candle throws his beams!
So shines a good deed in a naughty world.

Portia in *The Merchant of Venice*

What a 'groundwork' of moral philosophy can and cannot do

To avoid disappointment, readers of the *Groundwork* are well advised to keep in mind the very specific nature of Kant's project. What does he intend to achieve? Which questions does he not even try to address?

Let us start with what not to expect from a *Grundlegung*. The word can be used to describe the activity of laying the foundations of something or, when this is graced with success, its result.[1] We would therefore expect the *Groundwork of the Metaphysics of Morals* to contain the principles of another, distinct philosophical project. Indeed, that is how Kant describes the task that lies ahead in the Preface. The first five or so pages contain a careful discussion of the nature and necessity of a future metaphysics of morals, and it is only towards the end that Kant turns to the prior, 'critical' task of grounding this novel discipline. The slim volume is clearly part of the foundational project that preoccupied Kant during the 1770s and 1780s.

If the *Groundwork* does not claim to be a complete guide to ethical theory or moral life as a whole, it would be a mistake to try to reduce Kantian ethics to this book, even if we judge it to be Kant's most profound or influential contribution to moral philosophy.[2] The *Groundwork* leaves many questions to be addressed in later works, not just the new

1 Both Abbott's 'Fundamental Principles' and Beck's 'Foundations' fail to capture the subtly ambiguous nature of the German word. Ellington's 'Grounding' is arguably better than the now standard 'Groundwork'. Another possibility would be 'Foundation', in the singular.

2 Kant has a very acute sense of systematic priority. This is the reason why we should not infer from the fact that a certain topic is not again taken up in a later work that Kant has changed

'Metaphysics of Morals' itself. Kant occasionally mentions the idea that what we might call 'applied' ethics – moral psychology or 'anthropology', a project he never carried out – must follow a metaphysics of morality. There is also the idea of a fully fledged 'Critique' of pure practical reason, which settles fundamental problems that fall outside the narrow scope of the *Groundwork*. These projects are valuable in their own right, but they are not essential to the task of laying the foundations of ethics. We are not yet in a position to discuss, for instance, the comprehensive classification of first-order moral commands, the theological implications of the sum of all that is good, the unity of theoretical and practical reason, the casuistry of ethical conflict, the precise mechanism of moral motivation (and so forth). The purpose of the *Groundwork* is a more modest one: Kant 'merely' seeks to identify and firmly to establish the highest principle of moral volition.

There are also questions Kant does not intend to address in any of his works on moral philosophy. Most significantly, he does not intend to overthrow everyday morality. This is not due to some sentimental attachment to a particular moral code. Kant's reasons for thinking that common, pre-philosophical moral thought cannot be radically mistaken are largely ethical. Human beings must have access to moral truth to be responsible agents at all. Non-culpable ignorance renders attitudes of praise and censure invalid, and moral commands might be weakened to mere recommendations that apply on prudential grounds. If, as Kant argues throughout the *Groundwork*, moral action must be done 'for the sake of the law', all moral agents must have access to this law. He cannot dismiss common moral views and those who hold them as amoral. Kant is not immune to sceptical influences, but he takes some challenges – such as the twin threats of empiricism and physical determinism – more seriously than others. The *Groundwork* is not, therefore, an unbiased enquiry into what the grounding of morality might be, or whether there are moral principles at all. The truth of reflective common morality is the default position. Kant even makes what he considers the everyday notion of moral value the starting point of his enquiry: the only absolutely good thing is a good will (IV 393.5–7).

Kant's intentions have been misunderstood ever since the *Groundwork* was first published in 1785. In the preface of the *Critique of Practical*

his mind. For the most part, he simply moves on. Admittedly, in the course of his overall philosophical development the details of his system turn out to be rather more variable than he is usually prepared to concede.

Reason, he says that a critical reviewer of the *Groundwork* – in fact G. A. Tittel in his little commentary *Über Herrn Kant's Moralreform*, pp. 15–16 – 'hit the mark better than he himself may have intended when he said that no new principle of morality is set forth in it but only a *new formula*' (V 8 fn.). Kant rejects the idea of a novel principle as preposterous:

> For who would want to introduce a new principle of all morality and, as it were, first invent it? Just as if before him the world had been ignorant or in thoroughgoing error about what duty is. Whoever knows what a *formula* means to a mathematician, which determines quite precisely what is to be done to execute a task and does not let him miss it, will not take a formula that does this with respect to all duty in general as something that is insignificant and can be dispensed with. (V 8 fn.)

Moral truth, though universally accessible, is liable to be obscured by the all-too-human tendency to side with natural desire rather than reason. An explicit statement of the formula of morality can perhaps help to preserve its purity. As Kant puts it in the *Critique of Judgement*, an ordinary man and a – Kantian! – philosopher rely on the same rational principle when they judge, for example, fraud to be morally wrong, but the latter has a much clearer conception of it (V 228.18–20). Kant thought that the categorical imperative, particularly in the shape of its initial 'basic' formulation, can serve as a criterion or decision procedure in practical matters. He may have been too optimistic about the powers of common moral cognition or the educational potential of ethical theory, but it is important to note that the latter is a direct descendant of the former.

In Kant's ethics common understanding has a much more positive role to play than in his theoretical philosophy. In the *Metaphysics of Morals* of 1797 Kant goes so far as to christen sound reason an 'unwitting metaphysician' (VI 206.23). Pre-philosophical moral thought can get us started on the task of a metaphysics of morals, even if it cannot finish it. By contrast, Kant famously complains about the 'lazy method' (*bequemes Mittel*) of appealing to common understanding in speculative philosophy (*Prolegomena*, IV 259.12) because in that branch of human inquiry our natural prejudice in favour of sensibility obstructs the shift of perspective that is required for the metaphysics of nature to become a proper science: the 'Copernican turn' of the first *Critique*. It makes metaphysics of nature possible by allowing for a priori cognition of objects because the objects of knowledge themselves turn out to depend on our cognitive faculties (B xvi). Thus even if, by virtue of its foundational character, the latter parts of the *Groundwork* contain certain

philosophical technicalities and intricacies, it is still closer to everyday thought than, for instance, the *Critique of Pure Reason*, which at the time had been widely condemned as impenetrable (and in fact still is).

Pessimism and optimism in Kant's moral theory

We have already caught a glimpse of the striking combination of pessimism and optimism that confronts readers of the *Groundwork*. Kant is extremely confident about the cognitive and affective moral capacities of human beings, but he is also very sceptical about the actual moral quality of their conduct. Let us discuss these two convictions in turn.

The ease with which we are supposed to apprehend token moral truths stands in contrast with the endless complications of empirical cognition. For the latter, I must collect, arrange and process data, a procedure susceptible to all sorts of error. These uncertainties also affect instrumental – i.e. technical and especially prudential – reasoning, which makes use of empirical knowledge. By contrast, 'to see what I have to do in order that my volition be morally good', Kant declares towards the end of Section I, 'I do not need any far-reaching acuteness':

> Inexperienced with regard to the course of the world, incapable of being prepared for whatever might come to pass in it, I ask myself only: can you really will that your maxim become a universal law? (IV 403.18–22)

Of course, this is an early statement of the categorical imperative, the 'formula' of the supreme moral principle implicit in all ethical conduct.[3] Kant's standard example in this context is the question of whether one should return a deposit to the heirs of the rightful owner if one could easily keep it – he thinks it is clear that one should.[4] In the *Critique of Practical Reason* the case of the deposit is introduced by the assertion

3 The synthetic a priori nature of the moral principle – particularly its lack of reliance on uncertain empirical premises – accounts for Kant's optimism about moral cognition, at least in part. However, he seems to think that because he has ruled out one source of error he has excluded all. It would seem that even on the Kantian picture moral practice is complicated by the following factors: (i) the empirical effects of one's actions enter the moral equation even if the *principle* is synthetic and a priori and they do not therefore determine the result; (ii) arithmetic and geometry are also synthetic and a priori, but not everyone is a mathematical genius; (iii) instrumental reason is needed at the subordinate level of deciding how to put moral insight into practice, e.g. not whether but *how* to help. Ultimately, Kant's optimism is probably grounded in his conviction that moral commands must be categorical (universal and necessary) and in the egalitarian implications that follow. Complications (i) and (iii) are further discussed in Appendix D below.

4 See *Critique of Practical Reason*, V 27.21–28.3 and *Theory and Practice*, VIII 286.17–287.21. For an extensive discussion see my 'Depositum', *Zeitschrift für philosophische Forschung* 57 (2003), 589–600.

that even the commonest understanding can without instruction distinguish whether the form of a maxim is fit for universal legislation (V 27.21–2). It would seem that moral judgement works like the competent use of a natural language. Native speakers can effortlessly produce grammatical sentences and distinguish well-formed constructions from ill-formed ones. Yet they are unlikely to be aware of the principles they employ. Linguistic rules are made explicit only by subsequent philosophical analysis and reflection. What is more, bad theory is likely to corrupt language, as well as morals.

Kant's optimism is not confined to moral cognition. To defend the universal authority of the moral law he must also make sure that we have a motive at our disposal that is always sufficiently strong to produce the action we recognise to be right. After all, ought implies can. A standard cannot otherwise be categorical, i.e. independent of merely subjective motivational conditions.[5] At this point Kant makes some concessions to the frailty of human beings. We do not possess what he calls a 'perfect' or 'holy' will that effortlessly acts as the moral law bids. We merely possess 'pure' will – later called *Wille* as opposed to the faculty of choice or *Willkür* – which is governed by the laws that a metaphysics of morals must investigate. In short: not all, but part of our faculty of volition is pure. It is exposed to two forces: reason and inclination; and although in matters of conflict we must side with reason, we cannot make inclination go away. It can at best be conditioned to support our rational, especially moral projects. We can never become completely good. Reason is in essence the same in all of us, and so are the commands of morality, which depend on reason alone. Inclination, by contrast, displays huge variations. That is why moral action can be more or less difficult subjectively at different times, and why moral behaviour comes more easily to some human beings than to others. Kant is not blind to the diversity of humanity. He merely thinks that, for egalitarian reasons, human nature should not feature prominently in normative ethics.

In sum: the moral option is available to all agents, even if owing to natural inclination it cannot be the only option. As a result, all grown-up

5 This is a very delicate point. In Section III Kant has to admit that the existence of such a miraculous motive can merely be postulated, but not proved. We have reached the 'outermost boundary' of moral philosophy; see IV 459–63. Consequently, in the second *Critique* Kant confines himself to describing the workings of the moral incentive, but does not say how it is generated; see V 72.21–7. The moral interest in doing the right thing is identified with reverence for the moral law in a footnote at IV 399.40. Only actions motivated by reverence are morally good, see IV 440.5–7.

human beings capable of using their rational faculties are responsible for their moral failings. They *could* have behaved morally if they had chosen otherwise. In the second *Critique*, the voice of practical reason is said to make even 'the boldest offender' (*den kühnsten Frevler*) tremble with awe (V 80.1–2). In the *Groundwork*, Kant goes so far as to say that, when confronted with shining examples of virtuous conduct, 'the most hardened villain' (*der ärgste Bösewicht*) desires to be a moral man; and painful though it is for him to mend his ways, he *can* bring this about (IV 454.21–9).[6]

If all seems well on the prescriptive side of moral philosophy, Kant has much less faith in our capacity to detect the actual moral value of our actions. We *know* what we ought to do; we are convinced that we *can* act morally when moral action is required; owing to the influence of inclination we are rather less certain whether we *are going to* do it; and in retrospect, if on the face of it we have done the moral thing, we can never be sure that we did it for the right reasons and not for selfish ones, i.e. that the act was done *from* and not merely *in conformity with* duty. Kantian morality commands that we take the right attitude in action, not just the performance of the right act. An action is morally good only if it proceeds from a subjective principle or 'maxim' that is fit to be a universal law. But the moral quality of action remains obscure. We know human actions, our own and those of others, only as they appear to us in experience, and the regularities they display are part of the causal process of nature. Free actions do not surface as such.

However, Kant's scepticism about moral value extends beyond agnosticism. As inclination is dear to us in a way morality is not, he is very suspicious about the actual moral quality of people's character. Individual actions mostly coincide with duty, and everyone can be morally good, but very few of us actually are. Most human beings get their moral priorities wrong. This pessimism is manifest throughout the examples in Section I of the *Groundwork* (IV 397–9). Shopkeepers commonly treat their customers decently; but they do so because they care about their reputation, not on moral grounds. The anxious care people take to preserve or advance their lives is likely to be grounded in self-love, not in moral principle. Beneficent action is frequently the effect of our natural sympathetic tendencies, not ethical conviction. Kant argues that

6 The reason is that even he is endowed with a pure will; see explicitly Kant's lectures on ethics: *Collins*, XXVII 294.1. For a particularly clear statement of the thesis that it is always up to everyone to be moral (but not prudent) see *Critique of Practical Reason*, V 36.40–37.3.

any action motivated by inclination, rather than reverence for the moral law, lacks distinctive *moral* value, no matter how amiable the inclination in question may be. Moreover, human beings display a worrying tendency to conform the normative standards of action to their own desires, and to flatter themselves that they are more moral than in fact they are. They often wilfully mistake prudential regret for the nagging voice of conscience (*Collins*, XXVII 251.16–17). Innocence is easily corrupted – which is why moral philosophy is needed at all (IV 404–5). Unsurprisingly, non-sophisticated people seem to be rather better assessors of their own moral worth than clever intellectuals. Kant's spirit, as he puts it in the second *Critique*, 'bows before the common man' (V 77.1–5).[7]

Those who find the Kantian conception of moral value objectionably narrow should bear in mind that it is a consequence of his strong egalitarian convictions, which sentimentalism or virtue theory cannot accommodate. Kant considers it unrealistic to pretend that human beings can attain moral perfection, or that obligatory action is always pleasant or beneficial to the agent. These are elements of the human condition that a philosopher should not try to explain away. But we can at least presuppose that everyone is endowed with the same *capacity* to be moral, and create a level playing field in that respect. Morality must be about action, about what is up to us, not about the distribution of natural favours. There is a rather pithy handwritten note in which he says that 'it can be required of human beings that they act as the law commands; but not that they do so gladly' (R 8105, XIX 647). In the *Critique of Practical Reason*, Kant makes it quite clear that sympathetic feelings are often welcome, amiable, desirable, beautiful. They can under certain conditions be good objectively, all things considered. But they are not *morally* good (V 82.18–25). A happy, well-rounded character is an ideal that lies beyond the sphere of Kant's conception of morality.[8]

7 This was not always so. In a well-known handwritten note, Kant credits Rousseau with his democratic conversion: 'I am, in fact, a researcher by inclination. I feel the full thirst for knowledge and the unrest which goes with the desire to progress in it, as well as satisfaction at every acquisition. There was a time when I thought this alone could constitute the honour of humanity and I despised the ignorant rabble. Rousseau has set me right. This blinding prejudice disappears, I learn to honour human beings, and I would find myself more useless than the common worker if I did not believe that this kind of view can give worth to all others in establishing the rights of humanity.' See Kant's notes on his 1764 *Obervations on the Feeling of the Beautiful and the Sublime*, XX 44.8–16. Kant discovered Rousseau a year or two before the *Observations* were published.

8 Motivation and moral value are further discussed in Appendices A and B.

The character of moral duty

Kant's late work on the philosophy of religion contains some stunning examples of his confidence in common moral consciousness. In *Religion within the Limits of Reason Alone* he goes so far as to claim that conscientious moral judgement cannot err. The voice of conscience, which is our internal moral judge, can serve as a 'guiding thread' (*Leitfaden*) in matters of doubt. Kant introduces a new practical principle 'that does not stand in need of a proof': that we ought 'to venture nothing where there is a danger that it might be wrong'; in Pliny's Latin: *quod dubitas, ne feceris* (VI 185.24–5).[9]

To illustrate this point Kant focuses on the possibility of conflict between the commands of biblical faith and morality. His example is that of an inquisitor who condemns a decent citizen to death for his alleged heresy. Kant assumes that the death sentence is unjust. Yet the inquisitor did not make an innocent mistake. He 'consciously did wrong' because 'we can always tell him outright that in such a situation he could not have been entirely certain that he was not perhaps doing wrong' (VI 186.28–30). In other words, he is violating the new criterion of moral permissibility to use his sense of guilt as a guiding thread in moral matters, i.e. never to do anything which he is not completely certain is right. It is incumbent on the agent 'only to enlighten his *understanding* in the matter of what is or is not duty; but when it comes, or has come, to a deed, conscience speaks involuntarily and unavoidably' (VI 401.14–16).

Note that the inquisitor's conflict is practical but not moral. Kant is so confident that the inquisitor's sense of right and wrong could not be silent on the matter of whether he may kill an innocent person for religious reasons because he sees two distinct and quite unequal forces at work: revealed religion and pure practical reason. Kant does not question the sincerity of the inquisitor's faith; but no-one can ever be *certain*, as is morally required, that historical religion justifies the destruction of an innocent human being. Similarly, in the *Critique of Practical Reason*, the unconquerable voice of conscience is said to support our judgement that physical determinism does not suffice to undermine morality and responsibility (V 98.13–28). Again, the two opposing forces – nature

9 The idea that certainty can serve as a practical criterion is by no means an invention of the 1790s: moral 'probabilism' (VI 186.7) is rejected in the Methodology of the *Critique of Pure Reason*, where Kant argues that the mere opinion that an action is *permissible* is never sufficient to justify it (A 823/B 851). See also the much earlier reflection on moral certainty, R 2462 (XVI 380–1), as well as R 2504 (XVI 396).

and morality – are different in kind. Kant does not seem to envisage that we are torn between two courses of action for moral reasons. He makes no provisions for genuine moral dilemmas, where no option is unambiguously right or all options are equally problematic.

As the case of the inquisitor illustrates, Kant's conception of duty lacks many of the unpalatable connotations that the word might evoke in today's readers. It is important to keep this in mind. What human beings ought to do is not grounded in their social rank or station. After all, the moral law is universal. It is a result of one's status as a rational being amongst others. Moreover, in the last consequence we freely impose the law of duty upon ourselves – which is the definition of Kantian autonomy. In the late *Conflict of the Faculties*, Kant explicitly argues that the command of a superior is not valid automatically. To apply, it must be freely judged to be right (VII 27.27–30). It is a symptom of the perversity of National Socialism that Adolf Eichmann, in his Jerusalem trial, sought to justify his part in the slaughter of millions of Jews with reference to Kant's moral philosophy. Eichmann said that for a long time he was not just obeying orders, he was acting for the sake of the law – a law most certainly not his own – until finally he abdicated moral judgement to his superiors altogether.[10] But he could and should have known that he was *not* doing his duty.

A priori and a posteriori: the grounds of action

The opposition of the a posteriori (that which is grounded in the natural world of experience) and the a priori (the rational, which is not) pervades the whole of Kant's philosophy. Failure properly to distinguish between the two is just as pernicious in the practical realm as it is in the theoretical.[11] The opening statement of the 1787 edition of the *Critique of Pure Reason* (B 1) has close parallels in Kant's moral theory. In either field, sensibility comes first in the temporal order, but is not sufficient to accomplish the task at hand: the generation of knowledge or action. 'All our cognition', Kant states, 'begins with experience', and so do all our actions. For how else should our cognitive, or practical, capacities be 'awakened into activity [*Ausführung*]'? But it does not follow that either knowledge or action 'arises from' or is a mechanical product of

10 See H. Arendt, *Eichmann in Jerusalem* (Piper, 1986), pp. 231–5.

11 Note that for Kant, 'practical' does not have overtones of feasibility or 'practicality'. It is that which is concerned with action, rather than cognition. As reason on its own is 'practical' only in moral action Kant often uses the term synonymously with 'moral'.

empirical factors. In either case, the a priori addition of absolute spontaneity, i.e. activity of the self, is necessary to bring about the desired result.

This model of interaction between sensibility and reason is less conspicuous in the practical sphere than in the theoretical, but closer inspection reveals it to be a constant theme also in Kant's philosophy of action. For example, in the second *Critique* Kant maintains that inclinations 'always have the first word' (V 146.34, summarising V 74.8–15). This tendency can also be detected in the examples that Kant uses to illustrate the first variant of the categorical imperative in Section II of the *Groundwork*. The first illustration concerns an unfortunate man who is tired of life. Yet he is 'still so far in possession of his reason that he can ask himself whether it would not be contrary to his duty to himself to take his own life' (IV 421.25–422.3). Inclination does not automatically translate into action. He can still reflect on the courses of action open to him and decide in the light of rational considerations. In the second example, someone finds himself 'urged by need to borrow money' and he knows well 'that he will not be able to repay it; but he sees also that nothing will be lent him unless he promises firmly to repay it within a determinate time'. Kant says that this person 'is inclined' to make such a promise, but he has still 'enough conscience' to ask himself whether it is not perhaps 'forbidden and contrary to duty to help oneself out of need in such a way' (IV 422.15–20). The third and fourth examples, regarding our duties not to neglect our talents and to help those in need, follow the same pattern. Inclination has the first word but it need not – and often must not – have the last.

If sensibility is insufficient to produce either knowledge or action, something over and above the empirical is required to complete the process: they must both rest on principles that are subject to rational evaluation. In action, it is up to us to reject the pretensions of inclination. We can conform the subjective principles from which our actions proceed (maxims and rules) to objective principles of reason (imperatives). We would like to give in to our natural desires; but we are still free to do the right thing. Moreover, if knowledge and action are capable of rational justification, the grounds of theoretical and practical principles must be a priori. The a priori nature of ethical norms is borne out by the fact that, as in the case of knowledge, morality involves an element of necessity (see IV 389.11–13); but if Kant is right, necessity cannot be encountered in experience. Experience merely informs us about the way things are, not the way they ought to be (see B 3). Kant

takes the divide between (natural) is and (moral) ought very seriously indeed.

Kant's method: analytic, synthetic and the need for a 'deduction'

Kant is not exceptional in his professed reliance on commonly held moral beliefs to disclose their underlying principle. In the history of moral philosophy, even those who reach substantially different conclusions usually fail to come up with a plausible alternative. J. S. Mill is convinced that there is a 'tacit influence' on common moral judgement of an objective standard that can be revealed by philosophical means: the principle of utility (*Utilitarianism* I.4); and Aristotle's professed method in the *Nicomachean Ethics* consists in investigating reputable opinions or ἔνδοξα (Book VII, 1145b 2–7).[12] Starting with anything other than the views of common moral consciousness would expose Kant's project even more to the sceptical worry that his ethical theory is just a figment of a particularly lively philosophical imagination.

What is rather more remarkable than Kant's starting point is the division of labour assigned to the three sections of the *Groundwork*. Kant briefly discusses his research method towards the end of the Preface. Sections I and II are declared to proceed 'analytically', while Section III is said to be 'synthetic' (IV 392.17–22, cf. IV 445.7–8). In particular, the analytic sections are devoted to the identification or discovery of the highest principle of morality and its variations, which is then to be confirmed or justified in the final synthetic section (IV 392.3–4). Kant expects readers of the *Groundwork* to be familiar with the analytic/synthetic distinction from his theoretical writings (see IV 420.14–17). How is it employed in his moral philosophy?

The obvious reason why Kant considers the justification of morality to be in some sense synthetic is the following. A priori principles are either analytic or synthetic. The supreme principle of morality, if a priori, cannot rest on analytic foundations because the analysis of concepts helps us to *understand* them better but cannot establish their *reality* (see IV 420.18–23). Analytic judgements develop or clarify given concepts *without* assessing their validity – which is precisely what Kant does with the concept of duty in Section I and, in a more roundabout way, in

12 The question as to what extent Aristotle actually employs this method is disputed. For a critical assessment see J. Barnes, 'Aristotle and the Methods of Ethics', *Revue Internationale de Philosophie* 34 (1980), 490–511.

Section II. The concept of God by necessity points to his perfection or necessity; but this does not justify the assumption that there is a God. Analysing the concept of a 'bachelor' will reveal such a creature to be an eligible man who has never been married; but a woman interested in the existence and whereabouts of bachelors would be ill advised to confine her efforts to conceptual analysis.

That there are bachelors is an empirical judgement, unlike the equally synthetic judgement that God exists, or that human beings are subject to duties. As experience cannot vouch for our right to use concepts like God and duty, Kant is worried that our use of synthetic a priori principles may not be justified. In the case of some concepts it is rather doubtful whether they can be applied to reality at all, i.e. whether there is something that corresponds to them, and whether we say anything meaningful when we use them. In the *Critique of Pure Reason*, Kant mentions the examples of fortune and fate (*Glück, Schicksal*, A 84/B 117). His twelve categories are exposed to this suspicion because they are pure concepts of the understanding and as such not rooted in experience. To use Kant's term, borrowed from the legal literature of his day, synthetic a priori principles stand in need of a 'deduction':

> Jurists, when they speak of entitlements and claims, distinguish in a legal matter between the questions about what is lawful (*quid juris*) and that which concerns the fact (*quid facti*), and since they demand proof of both, they call the first – that which is to establish the entitlement or legal claim – the *deduction*. (A 84/B 116)

Consequently, Kant's deductions should not be confused with the standard 'top down' deductions or derivations of formal logic. They serve to corroborate that we are entitled to use a concept that can only be employed in a synthetic judgement a priori. For this purpose, we need to trace the origin of a concept and check whether the connection made is legitimate. This is one of the tasks indigenous to his critical philosophy. Section III, which provides as much of a 'Critique' as is needed for the purpose of grounding a moral metaphysics, is accordingly dubbed 'Transition from the Metaphysics of Morals to the Critique of Pure Practical Reason'.

In the last section of the *Groundwork*, Kant therefore intends to demonstrate, as far as is possible, that we are entitled to apply to human action the concept of duty as developed in the first and second sections. The principle of duty – the categorical imperative now in its final, most metaphysical variant: the principle of autonomy – must be both

synthetic and a priori. But like fate and fortune, the concept of duty might be no more than an 'empty concept' (IV 421.12), a natural and understandable idea to which nothing corresponds in reality. For all we know, human beings could be incapable of moral action, which would turn the analytic sections into the literally academic project of developing a fantastical concept.

This kind of philosophical concern should not be confused with the scepticism of the amoralist, who cannot see the point of moral action at all. Kant's problem is the worry of someone who is well disposed towards morality but cannot understand it. There is nothing in the world of experience – an otherwise reliable source of matching concepts and reality – that confirms the existence of duty. We can point to a bachelor when we see one, but we cannot point to a free human action done from a sense of duty, just as we cannot empirically identify an act of providence. Experience tells us about matters of fact, not norms, imperatives or values.[13] Yet if the world of experience cannot be the source of the authority of the moral law, what is? Or, as Kant himself puts it, 'whence' does the moral law obligate or 'bind' us (IV 450.16)?

There is now another sense in which the categorical imperative is a 'synthetic' practical principle (IV 420 fn.). Moral commands are characterised by the fact that the action commanded is independent of – not contained in or entailed by – any ends we want to pursue in action. A synthetic practical proposition tells the agent to do 'something new', just as a synthetic theoretical proposition provides us with new information beyond that which is contained in a specific concept (e.g. that a ball is blue, as opposed to round; or that the will is free, as opposed to a kind of causal power). By contrast, something in accordance with an analytic practical principle follows from a given end and does not constitute a separate action in its own right. When I add hot water to ground coffee beans to make coffee, I do not *both* add hot water *and* make coffee. According to the technical rules of coffee making, making coffee consists in, amongst other things, adding hot water to ground coffee beans. If you observe me as I pour water into a cafetière, my action can be explained with reference to an identifiable end or desire. The imperative I act on is hypothetical. If the human will was perfect like the will of God and unaffected by the obstacles put in its path by inclination, moral action would follow in a similar fashion. Unfortunately, we do not possess such a perfect will. The problem with human

13 Experience confined to things empirical – there is no 'moral experience' in this sense, despite the fact that the *Groundwork* commences with a notion of common moral understanding.

morality is that it does not rest on an antecedently given end that we wish to realise.

It is tempting to recast the problem of the possibility of synthetic practical principles in the terms of a prominent contemporary debate: how are external reasons possible? Or perhaps: how can my rational faculty create a new incentive not initially contained in my motivational set?[14] In slightly more old-fashioned terms: how can reason, just by itself, motivate? Kant's concept of duty makes sense only if it can.

The first *Critique* recognises the need to 'deduce' synthetic a priori judgements quite generally.[15] Even the concepts of space and time deserve a 'transcendental elucidation' – despite the fact that space and time, which form the ground of the synthetic a priori principles of arithmetic and geometry, are involved in experience. The two concepts rest on pure forms of intuition and are therefore not empirical in origin (B 40–1, B 48–9). The twelve pure concepts of the understanding or 'categories', the application of which makes experience possible in the first place, require a fully fledged transcendental deduction, even though they are at least indirectly confirmed by experience (A 84 ff./B 116 ff.). If so, it should be obvious why the position of moral concepts like 'duty' and 'autonomy' is so precarious. They are a priori in origin, but they have no possible link with experience at all. They are not even, like the categories, capable of indirect corroboration; and what is worse, experience appears to confirm that all human action is subject to natural laws and therefore, by definition, *not* free.

The parallel between the theoretical and the practical that was introduced in the previous section is now complete. Kant resumes the argument quoted above as follows. There remains, he says,

> a question which at least requires closer investigation, and one not to be dismissed at first glance, whether there is any such cognition independent of experience and even of all impressions of the senses. One calls such cognitions

14 See B. A. O. Williams, 'Internal and External Reasons', in *Moral Luck* (Cambridge University Press, 1981). However, Kant's question cannot easily be mapped on to the modern debate. The categorical imperative, as a command of reason, is like an 'external' reason in that it commands independently of the agent's current motivational state. But, for Kant, reason can independently cognise an action to be right and even motivate us to act. The 'reasons' that emerge would be *external* to the agent's initial motivational set, but *internal* to the agent. Kantian autonomy is wholly opposed to the normative authority of any so-called reason that is externally imposed upon the agent. Moreover, Kant's notion of a moral imperative is stronger than the standard modern notion of a moral reason in that it provides reasons that are not just motivating or overriding, but even necessary to the exclusion of all other reasons.

15 See A 232–3/B 285–6, contra the alleged self-evidence of principles, and A 149/B 189.

a priori, and distinguishes them from *empirical* ones, which have their sources *a posteriori*, namely in experience. (B 1–2)

In like manner, any critical investigation of practical reason will have to investigate the question whether there are any *actions* completely independent of everything empirical, and their sources a priori. In Section III of the *Groundwork*, the question of how much *pure* reason can by itself accomplish – in particular: how synthetic practical principles a priori are possible – once again defines the project of a (rudimentary) second 'Critique' of pure reason.

The story of the Groundwork

When the *Groundwork* appeared in print in the spring of 1785, it was Kant's first published work devoted exclusively to the subject of moral philosophy. But Kant was not a newcomer to the discipline.[16] By the mid-1780s, he had been planning to write a book on the foundations of ethics, entitled 'Metaphysical Principles of Practical Philosophy', 'Critique of Moral Taste' or 'Metaphysics of Morals', for at least twenty years.[17] In February 1767, J. G. Hamann told Herder that Kant was working on a 'Metaphysics of Morality', which unlike previous ethical theories was meant to investigate the question of 'what man is, rather than what he ought to be' (IV 624); and on 9 May 1768 Kant wrote to Herder expressing his hope that he might complete a 'Metaphysics of Morals' by the end of that year (X 74, No. 40 [38]). But, like his theoretical philosophy, Kant's ethical theory soon changed beyond recognition. He abandoned the idea that a metaphysics of morals should be a descriptive, psychological study of human nature. By the early 1770s, the 'first grounds' or 'pure principles' of morality had become part of the new critical project of exploring 'The Boundaries of Sensibility and Reason', published as the *Critique of Pure Reason* a decade later.[18] In the *Groundwork*, Kant denounces the empiricist project of moral enquiry as

16 On the composition of the *Groundwork* cf. P. Menzer's introduction to the Academy text of the *Groundwork*, IV 623–9; B. Kraft and D. Schönecker's introduction to their edition, VII–XIII; Manfred Kuehn, *Kant. A Biography* (Cambridge University Press, 2001), pp. 277–328.

17 Kant mentions the first title as that of a treatise on moral philosophy in a letter to J. H. Lambert dated 31 December 1765, X 54–7, No. 34 [32]; it was as announced in Kanter's 1765 Michaelmas publisher's catalogue under the second title, but never appeared in print (see Menzer's introduction, IV 624 fn.).

18 See the letters to Marcus Herz dated 7 June 1771 (X 121–4, No. 67 [62]) and 21 February 1772 (X 129–35, No. 70 [65]).

at best irrelevant and at worst pernicious. Pure normative moral theory must precede moral psychology or 'anthropology'.

The idea that a dual metaphysics of morals and nature should, in that order, follow the critical foundations of transcendental philosophy remained a constant feature of Kant's philosophical ambitions. It is first mentioned in a letter to Marcus Herz in late 1773 (X 145.20–2, No. 79 [71]). However, on more than one occasion Kant changed his mind as to how much of a 'critical' preparation was needed to ground the moral part of the metaphysical system. Moral philosophy was very much part of the critical enterprise in the early 1770s but was then discarded, either because Kant realised that he had enough work on his hands with laying the foundations of the metaphysics of nature, or because he thought that the later *Critique of Pure Reason* provided a sufficient foundation of both parts of the twofold metaphysics. At the time of composition of the Methodology (see A 841/B 869), Kant did not seem to feel the need of a grounding of metaphysics, moral or natural, other than the 1781 *Critique* itself, which after all makes room for freedom and responsibility.[19] Yet a rudimentary *Critique of Pure Practical Reason* was published in 1785 under the title of a *Groundwork of the Metaphysics of Morals*. The second edition of the *Critique of Pure Reason*, which for a while was meant to revert to the original plan and cover moral as well as speculative philosophy, followed in 1787; the *Critique of Practical* [sic!] *Reason* as an independent publication in 1788; and by 1790, Kant deemed an additional third 'Critique' necessary to complete the foundation of the dual metaphysical system: the *Critique of Judgement*.[20]

The composition of the *Groundwork of the Metaphysics of Morals* is shrouded in mystery. As Hamann reveals in a letter to Kant's publisher, J. Fr. Hartknoch, in January 1782, Kant returned to working on a 'Metaphysics of Morals' soon after the publication of the *Critique* in the previous year (IV 625). However, we possess some evidence to the effect that Kant's intention to 'issue [a] groundwork in advance' (IV 391.17) was influenced by the publication of Christian Garve's

19 In the Introduction, Kant excludes moral matters from transcendental philosophy because 'for that, the concepts of pleasure and displeasure, of desires and inclinations, of the faculty of choice etc., which are all of empirical origin, would have to be presupposed [*vorausgesetzt werden müßten*]' (A 14–15, see A 801/B 829 fn.). (The claim is weakened in the second edition: empirical concepts are no longer presupposed but still 'drawn into' moral philosophy; see B 28–9. H. Vaihinger detects in this, and the inclusion of aesthetics at B 36 fn., signs of an incipient broadening of the scope of the critical project; see his *Commentar zu Kants Kritik der reinen Vernunft* (W. Spemann, 1881), vol. I, p. 483.) However, on this early conception the *Critique* is still philosophically prior to moral philosophy by making room for transcendental freedom; see A 805/B 833.

20 Unlike the other two *Critiques*, the *Critique of Judgement* lacks a corresponding metaphysical doctrine; see V 170.20–7, V 168.30–7.

annotated German translation of Cicero's *De officiis* in 1783. Kant had held Garve, a 'popular philosopher' at Leipzig, in high regard. He was hoping to recruit Garve for the critical cause and was therefore disappointed to learn that he was the author of a scathing anonymous review of the first *Critique* published in the influential *Göttingische Anzeigen von gelehrten Sachen* in January 1782. It had, admittedly, been abridged, edited and not at all improved by J. G. H. Feder, who was professor of philosophy at Göttingen. The *Prolegomena* represents Kant's reply. When after a conciliatory exchange of letters between Garve and Kant (X 328–33, No. 201 [184], and X 336–43, No. 205 [187]) the original review was published in 1783 in the *Allgemeine deutsche Bibliothek*, Kant still had little reason to be impressed. It was again Hamann who, in a letter to Scheffner in February 1784, reported that Kant was working on a 'Counter-Critique' (*Antikritik*) of Garve's 'Cicero' that was, as a matter of fact, intended as a retort against the unabridged review of the *Critique* (IV 626). It is difficult to say whether Hamann's testimony is credible.

Kant would have been upset by Garve's new publication – even if the two men had not previously come into conflict – by Garve's blatant, uncritical eudaemonism as well as the lack of systematic rigour of his supplementary *Philosophical Remarks and Treatises*, rather than his translation of Cicero's three books *On Duties*, with which Kant had long been familiar in the original Latin.[21] In other words: if it is true that 'Garve's "Cicero"' inspired Kant to turn his attention to the foundations of moral philosophy it was probably Garve's work, rather than Cicero's. Yet by the end of April 1784 Kant apparently decided to abandon the plan of writing a response to Garve in favour of a short, foundational ethical treatise – a *prodromus* or 'forerunner' of moral philosophy, as the ever-prolific Hamann calls it in his letters (IV 627). If for a while Kant still intended to attach a direct reply to Garve as an appendix to the *Groundwork* it did not find its way into the final version.[22] Kant sent the manuscript of the *Grundlegung* to Hartknoch in September 1784. It was published, after some delay at the printer's office, at the Easter book fair of 1785. Kant received his first copies on 8 April of that year.

21 Apparently, Kant did not hold Cicero in particularly high regard. In the *Conflict of the Faculties*, he recommends repeating Cicero's name to oneself in bed as a soporific – the philosophical equivalent of counting sheep (VII 107.2–3). His disappointment with Garve must have prevented him from using his name to even better effect.

22 See Hamann's letter to Lindner, dated 9 March 1785 (IV 628). Kant finally directly attacked Garve's moral philosophy, or rather his lack of understanding of the *Critique of Practical Reason*, in Section I of the essay on *Theory and Practice* in 1793.

It is difficult to say how much of Kant's response to Garve's 'Cicero' was in the end incorporated into the *Groundwork*.[23] The fact that Kant mentions neither philosopher should not, of course, be taken as evidence that he did not intend them to be his targets – Kant rarely refers to his most prominent opponents by name, and he may well have thought that he succeeded in elucidating common moral thought while Garve, who pretends to do the same on Cicero's behalf, failed.[24] Moreover, there are some striking similarities between the two projects. At a rather superficial level, Cicero divided his brief treatise on duty into three sections (or 'books'), and so did Kant. More interestingly, the *Groundwork* – particularly Section I – contains a plethora of allusions to ancient themes: Kant rejects an ethics of social status, the moral sufficiency of a desire for honour and, above all, the identification of happiness with the highest good. The first variant of the categorical imperative – the formula of universal laws of nature – is clearly intended as a sensible restatement of the Stoic thesis that we should strive to live in accordance with nature. But this is where similarities end. Kant did not need Garve's translation to remind him of the Stoic principle, which was still popular with eighteenth-century thinkers like Wolff and Baumgarten; and the other variants are hardly directed against Cicero. Moreover, the *Groundwork* is too complex, even as a piece of philosophical rhetoric, to be inspired by two second-rate philosophers. Kant adapted the notion of a moral commonwealth or 'kingdom of ends' from Leibniz, and it had been in place as an ideal long before Kant wrote the *Groundwork*. Most importantly, the main innovation of the *Groundwork*, Kant's theory of morality as autonomy, can hardly be reduced to a reaction to Garve.[25]

23 See K. Reich's *Kant und die Ethik der Griechen* (Mohr, 1935) and, more recently, Carlos Melchios Gilbert, *Der Einfluß von Christian Garves Übersetzung Ciceros 'De Officiis' auf Kants 'Grundlegung zur Metaphysik der Sitten'* (S. Röderer, 1994) for rather too positive accounts of Garve's, or Cicero's, influence. The influence of Reich can also be felt in the commentaries of A. R. C. Duncan and J. Freudiger, who toy with the idea of bracketing the variations of the categorical imperative in Section II as a mere rhetorical interlude. For a more balanced discussion see D. Schönecker, *Kant: Grundlegung III. Die Deduktion des kategorischen Imperativs* (Alber, 1999), pp. 61–7, and M. Kuehn, 'Kant and Cicero', in *Kant und der Berliner Aufklärung*, ed. V. Gerhardt, R.-P. Horstmann and R. Schumacher (De Gruyter, 2001).

24 See Christian Garve, *Philosophische Anmerkungen und Abhandlungen zu Ciceros Büchern von den Pflichten* (Wilhelm Gottlieb Korn, 1783), vol. III, pp. 262–3.

25 A closely related change that a reaction to Cicero cannot account for is Kant's complete exclusion of God and religion from the foundations of ethics. In the lectures on moral philosophy, religion was needed to guarantee the existence of an interest in doing the moral thing (see *Collins*, XXVII 308–10); and we have a duty to God to comply with our obligations from duty (see XXVII 272.4–8). With the benefit of hindsight we realise that this position is unstable. Kant had to abandon this rather uneasy division of incentive and determining ground. The law of morality must be our very own command.

Kant's Groundwork*: synopsis of the argument*

Preface

1 Classification of the disciplines of philosophy, according to their subject matter and mode of cognition (IV 387–8)
2 Why pure moral philosophy, or a 'metaphysics of morals', is necessary (IV 388–91)
3 The project of *grounding* a metaphysics of morals (IV 391–2)

Section I: Transition from common to philosophic moral cognition of reason

1 On the unconditional value of a good will (IV 393–4)
2 A morally good will, not happiness, is the natural purpose of reason (IV 394–6)
3 Elucidation of the concept of duty by means of three propositions (IV 397–401)
   a The general concept of a good will must be made more determinate by analysing the concept of duty (IV 397)
   b Proposition 1: An action that coincides with duty has moral worth if and only if its maxim produces it by necessity, even without or contrary to inclination (IV 397–9)
   c Proposition 2: The moral worth of an action does not lie in the effect intended but rather in its maxim [to be judged by the standard of a formal principle] (IV 399–400)
   d Proposition 3: Duty is the necessity of an action from reverence for the law (IV 400–1)
4 The law of duty, general conformity to law as such, is the condition of a will that is good in itself (IV 402–3)
5 Concluding remarks: common and philosophic moral cognition of reason (IV 403–5)

Section II: Transition from popular moral philosophy to the metaphysics of morals

1 Preliminaries (IV 406–12)
  a The origin of the concept of duty is not empirical but a priori (IV 406–8)
  b On the limited value of exemplars in ethics (IV 408–9)
  c True and false popularity in moral philosophy (IV 409–10)
  d The primacy of metaphysics in moral philosophy (IV 410–12)
2 The doctrine of imperatives (IV 412–20)
  a The will as the capacity to act in accordance with the representation of laws (IV 412–13)
  b Imperatives necessitate an imperfect will to act in accordance with laws (IV 413–14)
  c Imperatives, hypothetical and categorical: skill, prudence, morals (IV 414–17)
  d How are all of these imperatives possible? (IV 417–20)
3 The categorical imperative (IV 420–1)
  a Derivation of the general formula of the categorical imperative from its concept (IV 420–1)
  b The general formulation (IV 421)
4 The first variant: universal laws of nature (IV 421–4)
  a The universal-law-of-nature formulation (IV 421)
  b Application of this formula to the four examples of duty (IV 421–4)
5 Interlude (IV 425–7)
6 The second variant: rational creatures as ends-in-themselves (IV 427–31)
  a Derivation of the 'formula of humanity as the end-in-itself' from the concept of a will (IV 427–9)
  b Application of this formula to the four examples of duty (IV 429–31)
7 The third variant: autonomy in a kingdom of ends (IV 431–6)
  a Derivation of the formula of autonomy from the other two (IV 431)
  b A universally legislative will is independent of all interest (IV 431–3)
  c Self-legislation, morality and the kingdom of ends (IV 433–4)
  d A moral being possesses dignity, not a price (IV 434–6)

8 Reflections on the variant formulations of the categorical imperative (IV 436–40)
  a The connection between the three variants of the categorical imperative (IV 436–7)
  b Review of the *Groundwork* so far: the good will and the formulations of the categorical imperative (IV 437–40)
9 The autonomy of the moral will (IV 440–4)
  a Autonomy and heteronomy (IV 440–1)
  b Division of ethical theories according to the principle of heteronomy (IV 441–4)
10 Transition to Section III: how is a synthetic practical proposition possible? (IV 444–5)

Section III: Transition from the Metaphysics of Morals to the Critique of Pure Practical Reason

1 The concept of freedom is the key to the explanation of the autonomy of the will (IV 446–7)
2 Freedom as property of the will of all rational beings (IV 447–8)
3 The interest attaching to the ideas of morality (IV 448–53)
  a Preparation of the 'circle': our consciousness of freedom and morality are not grounded in any conventional interest (IV 448–50)
  b The suspicion of a 'circle': freedom and morality (IV 450)
  c The escape: we step outside the circle when we consider ourselves members of an intellectual world (IV 450–3)
4 The 'deduction': how is a categorical imperative possible? (IV 453–5)
5 The extreme boundary of all practical philosophy (IV 455–63)
  a The problem of reconciling natural necessity and free will does not yet mark the extreme boundary of practical philosophy (IV 455–7)
  b We are conscious of our free will but cannot cognise or explain it (IV 457–9)
  c The inexplicability of the interest we take in morality is the outermost boundary of moral philosophy (IV 459–63)
6 Conclusion: Comprehending that we cannot comprehend morality (IV 463)

Commentary

The Preface

The Preface approaches the task of grounding the novel philosophical discipline of a 'metaphysics of morals' in three consecutive steps. First, taking his cue from the tripartite ancient division of philosophy into physics, ethics and logic, Kant systematically maps out the various philosophical disciplines and then directs our attention to the part of pure philosophy called 'metaphysics'. He secondly restricts his focus to pure moral philosophy, i.e. a metaphysics of morals as opposed to the more familiar metaphysics of nature, and emphasises its supreme importance and practical relevance. Thirdly and finally, Kant turns to the specific task and method of the present project: laying the foundations of a metaphysics of morals. He declares that he intends to pursue the project of such a metaphysics at a later date.

1 Classification of the disciplines of philosophy, according to their subject matter and mode of cognition

¶ IV 387.2 The first pages of the Preface reflect Kant's belief that pioneering work in a particular philosophical discipline should be preceded by locating it in the system of philosophical enquiry as a whole. For this purpose, he turns to the classical division of philosophy into physics, ethics and logic. It is commonly attributed to Xenocrates (396–314 BC), the third head of Plato's Academy, and was widely accepted in later antiquity, particularly in Stoicism.[1] Kant thinks that the ancients discovered this classification in a somewhat haphazard fashion, but he does not object to the trichotomy as such. The tripartite division is perfectly

1 See A. A. Long and D. N. Sedley, *The Hellenistic Philosophers* (Cambridge University Press, 1987), vol. I, pp. 158–62 and vol. II, pp. 163–6. Kant discusses the tripartite division in a similar fashion in his 1784–5 lectures on practical philosophy: *Mrongovius* II, XXIX 630; it is reaffirmed in very similar terms in the *Critique of Judgement*, V 171–2.

reasonable, even if the ancient authors were unaware of its underlying principles. It is the first task of the Preface to rectify this matter. Kant wants to make sure that the classification is 'complete', i.e. that no part of philosophy is missing; and he wants to be able to discern certain 'necessary subdivisions' within the three disciplines, such as metaphysics.

That human reason at first articulates its principles imperfectly is a well-known Kantian theme. In the *Critique of Pure Reason*, Kant famously attacks Aristotle for having 'rhapsodically' – and only partially – assembled his categories (A 80/B 106–07); in the *Conflict of the Faculties*, the idea of a university is traced to the implicitly rational plan of a single person who first suggested the foundation of such an institution to the government of his day (VII 17.2–17);[2] and, as we shall see below, the variant formulations of the categorical imperative in Section II of the *Groundwork* are an attempt to preserve what little truth is contained in the flawed ethical theories of Kant's philosophical rivals.[3]

¶ IV 387.8 Kant's attempt to re-establish the ancient division of philosophy proceeds as follows. Logic is said to be purely formal because it does not as such apply to any specific subject matter: the laws of logic are valid irrespective of the objects one happens to think about.[4] Yet there are also philosophical disciplines that by their very nature refer to certain objects: physics and ethics. These objects are characterised by, and therefore cognised through, their own proper laws. Kant now tacitly introduces the contentious assumption that there are precisely two kinds of such laws: laws of nature in the domain of physics, and laws of freedom in the domain of ethics. Both kinds of laws are causal laws.[5]

If we (i) accept the philosophical distinction between form and matter, and moreover share Kant's assumptions that (ii) all formal philosophy in the above sense is logic, and that (iii) all material philosophy requires certain laws, of which there can be only these two varieties, we are led to share Kant's conclusion that the ancient tripartite division

2 See also VII 21.5–21 and Kant's preliminary sketches of the introduction to the *Conflict*, XXIII 429–30.

3 For a more systematic account of the 'seeds' that lie hidden within reason see the Architectonic of the first *Critique*, A 832–5/B 860–3.

4 Nor do the formal laws of logic select any particular matter, unlike the equally formal moral law.

5 He returns to the details at IV 446.7 below. See also *Critique of Pure Reason* A 532/B 560 on 'two kinds of causality', and quite explicitly R 7018, probably 1776–8: 'All laws are either [laws] of nature or of freedom' (XIX 227).

of philosophy as either logical, physical or ethical is correct as well as complete.[6]

¶ IV 387.17 It is impossible for logic, Kant now argues, to have an empirical part. The reason given is that logic must already be in place as a strictly universal and necessary standard or 'canon' (*Kanon*) of all thought for us to engage in empirical research. Empirical investigation *presupposes* logic. It is rather surprising, then, that a little later Kant himself speaks of pure as well as 'applied' logic (and mathematics, IV 410, fn.), all the more so as he explicitly mentions an alleged analogy with a *pure* metaphysics of morals and an *applied* moral 'anthropology' – the eighteenth-century term for the discipline now better known as 'psychology'. Nor does Kant's discussion of general and particular, as well as pure and applied, logic in the *Critique of Pure Reason* fit the present picture (A 52–5/B 76–9). There seems to be an empirical part of logic – an essentially psychological investigation of the logical function and malfunction of reason – that at IV 387.17 is arbitrarily excluded from philosophy.[7]

Could we solve the problem by rejecting the implicit equation of the 'empirical' and the 'applied'? If so, logic may have an applied but not an empirical part. This strategy falters because on Kant's conception 'applied ethics' no more depends on empirical matters than does 'applied logic' and any possible distinction between what Kant calls the 'empirical' and the 'applied' soon disappears again (see e.g. IV 389). A more promising way to resolve the difficulty, or at least reduce the tension, might be the following consequence of the formal/material distinction. In the Preface, Kant is hinting at the fact that 'applied' logic is no longer part of formal logic; and only formal logic strictly qualifies as a 'science', and as the standard of all thought (see A 54/B 78). Logic does not have a proper field of application, with its own material laws, and it is not in any way enriched or augmented by being applied. By contrast, the application of metaphysics – both of nature and of freedom – is still part of, and makes a genuine contribution to, the broad

6 Note that these disciplines are 'doctrines', i.e. secure bodies of knowledge that first require – at least in the case of physics and ethics – some 'critical' preparation. See B ix–x.

7 It is worth noting that Kant was less convinced of the purity of moral philosophy when he wrote the first *Critique* in 1781; see the changes made to the text at A 15/B 29. However, the parallel treatment of pure logic/moral philosophy and applied logic/a doctrine of virtue at A 55/B 79 is left unchanged. Note also that the characterisation of the latter discipline as concerning 'the impediments of feelings, inclinations and passions' is very similar to that of a 'moral anthropology' at the end of the present paragraph (IV 388.1–3).

disciplines of physics and ethics. They must ascertain the laws pertaining to their proper objects: descriptive laws of experience (of that which *exists*) in the first case, normative laws of morality (of what in the case of human volition, which is potentially sidetracked by non-rational influences, *ought to be*) in the second (see Jaesche's *Logic*, IX 18). The more empirical part of ethics is also concerned with the factors that prevent human beings from acting in compliance with moral laws.

¶ IV 388.4 With a view to potential sub-disciplines within the tripartite scheme Kant now explicitly distinguishes 'pure' and 'empirical' philosophy. Pure philosophy concerned – unlike formal logic – with some specific object is called 'metaphysics'. In a general Kantian sense, metaphysics is the systematic a priori investigation of the most fundamental laws that govern cognition and action.

¶ IV 388.9 Corresponding to the traditional 'metaphysics of nature' there is now the notion of a 'metaphysics of morals' (see A 841/B 869). Kant thus broadens the 'literal' meaning of metaphysics – the continuation of physics with different means – to extend to the practical sphere.[8] In the first *Critique*, Kant says that metaphysics is needed 'not for the sake of natural science but instead to get beyond physics' (B 395 fn.). That is why the metaphysics of nature has 'especially appropriated' the name of metaphysics, as Kant writes in the Methodology (A 845/B 873). It is the kind of metaphysics that forms the subject of the (preparatory) *Critique of Pure Reason* and the *Metaphysical Foundations of Natural Science*. The metaphysics of nature must be distinguished from the 'empirical doctrine of nature', the bulk of physics in our modern sense of the word.

Calling the empirical part of ethics a 'practical anthropology', Kant is not referring to the *Anthropology from a Pragmatic Point of View* of 1798. As we are to learn later, the term 'pragmatic' refers to human 'welfare' (IV 417.1), i.e. to something decidedly non-moral, whereas 'practical', at least in a narrow sense, is synonymous with 'moral'.[9] Rather, a *practical* anthropology would consist in a detailed account of human moral psychology with all its failings and shortcomings, a subordinate project

8 The term 'metaphysics' originally referred to Aristotle's writings on ontology, which followed the treatises on physics in Andronicus' arrangement of the Aristotelian corpus. It was then, in a figurative sense, applied to the investigation of the mysteries that lie beyond natural science.

9 Cf. Kant's attempt to reclaim the word 'practical' in the second *Critique*, V 26 fn.

that Kant never executed. In his *oeuvre* as we read it today, the more applied remarks of the ethical second part of the *Metaphysics of Morals* are probably closest in spirit to a 'practical anthropology'.

2 Why pure moral philosophy, or a 'metaphysics of morals', is necessary

¶ IV 388.15 Kant has now assigned ethics and physics their proper place in the system of philosophical disciplines. He has also determined their sub-disciplines, pure and applied; which raises the practical questions of how they are related. With a side-swipe at the populist moral philosophers targeted in Section II, Kant brackets the more ambitious question of whether 'pure philosophy' calls for the professional care of a specialist, in analogy with the division of labour that in the wake of Adam Smith's work was beginning to be so fruitfully employed in many other areas of human activity in the late eighteenth century; but he proceeds to argue that in natural as well as moral philosophy the pure part must be sharply separated from the empirical.[10] We thus arrive at the twofold task distinctive of Kant's critical philosophy: what can pure reason achieve in these two 'material' disciplines? What are their a priori sources?

¶ IV 389.5 Finally leaving the metaphysics of nature behind, Kant confines the subject to morality. Should we not, he inquires, 'work out for once a pure moral philosophy, completely cleansed of everything that may be purely empirical and that belongs to anthropology' (IV 389.6–9)? The need for such an ethical theory is said to follow from generally held moral ideas. What does his argument for the need of a pure moral theory look like?

Kant considers it to be obvious – in fact part of the meaning of the very term – that 'moral laws' (if they exist at all) must be strictly necessary and universal. We are not allowed to exempt ourselves from the general command not to lie if we judge a truthful declaration to be inconvenient.[11] The reason is not to be sought in our specific constitution as human beings. We are subject to this unconditional command by virtue of our capacity to let action be guided by reason. That is why

10 Kant generally had a clear sense of the separateness of different types of academic enquiry. See e.g. *Critique of Pure Reason* B viii and A 842/B 870, *Conflict of the Faculties* VII 7.

11 None of this prejudges the question of when exactly a command applies. Kant is merely emphasising that compliance must not depend on subjective conditions.

a prohibition of lying does not just apply to all human beings alike, but also to all other rational beings similarly capable of insincerity. According to Kant, strictly universal and unconditionally necessary moral laws cannot be grounded in experience because experience at best informs us about facts, about the way things are. It teaches us contingent truths only.[12] If so, there can be no such thing as a proposition that is both necessary and a posteriori. Kant concludes that an ethical theory grounded in experience cannot account for the rigour that defines moral laws.[13]

Is it legitimate to infer the possibility, or even necessity, of a pure ethical theory from the impossibility of an ethical theory founded on empirical principles, as seems to be Kant's strategy in the Preface? Is it not conceivable that we are incapable of investigating the conditions of absolute moral commands, even if we suspect that they are rooted in reason? On the last pages of the *Groundwork* we discover that Kant takes the latter, more sceptical view (see IV 455–63). For the time being, however, his optimism about the power of reason serves to close the gap. It is not absurd to assume that the commands of practical reason are fit to be the subject of rational enquiry.

The present paragraph presents us with further difficulties. First, Kant's favourite example of a moral command may appear problematic because it seems to introduce into a metaphysics of morals a concept apparently borrowed from experience: the concept of lying. Secondly, the universality of the prohibition to lie might be called into question. It does not appear to apply to rational beings who cannot communicate with each other, or to those constitutionally incapable of being untruthful.[14] Neither objection is convincing. To begin with, Kant is not collecting empirical data. As in Section I of the *Groundwork*, he is developing his moral philosophy on the basis of reflective normative convictions that are generally shared, at least implicitly, and that he considers to be essentially correct. The purpose of this procedure is wholly heuristic. This method is entirely legitimate in a foundational work on moral philosophy. In fact, it is difficult to see an alternative.

12 See e.g. *Critique of Pure Reason* A 91/B 123–4 and B 142.

13 In his lectures on moral philosophy, Kant similarly argues for the aprioricity of the supreme principle of morality on the grounds that other theories fail to account for its unconditional nature. He rejects egoist, sentimentalist and social reconstructions of the command not to lie. For instance – contra moral sense theories – if there was 'anyone not possessed of a feeling so fine as to produce in him an aversion to lying' he 'would be permitted to lie' (*Collins*, XXVII 254).

14 See the *Anthropology*, for such a scenario (VII 332.13–21); or Lucian's example of the Greek god Momus, quoted by Kant in his lectures, who wanted 'that Jupiter should have installed a window in the heart so that every man's fundamental attitude [*Gesinnung*] might be known', which is thought to lead to the general improvement of people's moral principles (*Collins*, XXVII 445).

The presupposition of common moral beliefs need not infect the purity of the final product, the metaphysics of morals.[15]

Similarly, Kant is not committed to the view that the command not to lie as such is part of a pure moral philosophy, though it may be a consequence. Rather, he intends to argue that our consciousness of the strictness of this command points to the non-empirical origin of the principles of morality. Moral philosophy must therefore proceed by means of a priori reasoning. Furthermore, Kant need not hold the view that the command not to lie must apply to all rational creatures. It is obvious right from the start that qua command it fails to apply to a rational being endowed with a 'holy' will because such a perfect being does not need to obey moral commands: like the mythical Balliol man, it is effortlessly superior. The universal validity of the prohibition implies that any rational creature who, like us, faces the decision of whether to lie or to speak truthfully is subject to it. In fact, our moral intuitions confirm this. Consider our reactive attitudes towards the fantastic alien creatures we encounter in science fiction. Extraterrestrial beings who destroy planet Earth, eradicate humanity or at the very least desire to eat our cats are different in kind from natural disasters even if the effects are indistinguishable. Perhaps the only thing they share with us is the use of reason and language. Yet we naturally judge them by the moral categories of good and evil.[16]

¶ IV 389.24 We are now in a position to see why a separate metaphysics of morals is not just a worthwhile philosophical pursuit but even a necessity. The pure part of ethics is primary in every respect. Not only should we separate the empirical from the pure part of moral philosophy; the former must be subordinated to the latter, just as in cases of conflict the requirements of practical reason take precedence over the sensual needs of human beings.

Experience still has a twofold role to play. First, applying moral commands to specific cases requires experience in a somewhat different sense: practice. We need to learn to decide whether a moral command

15 See Kant's warning at the beginning of Section II, IV 406.5–407.16.

16 Insisting that pure moral philosophy must precede empirical moral philosophy or psychology Kant distances himself from his earlier position, which was influenced by the moral sense theories of Shaftesbury, Hutcheson and Hume. Announcing his lectures in the winter semester of 1765–6 he states that, historically and philosophically, that which happens must be considered before progressing to that which ought to happen; and that 'the nature of man' (*die Natur des Menschen*) must be studied first (II 311.31–2). The main lectures on moral philosophy – with their ambiguous theory of moral motivation – seem to occupy an intermediate position: see M. Kuehn's introduction to W. Stark's new edition of *Immanuel Kant. Vorlesung zur Moralphilosophie* (De Gruyter, 2004), p. XXVIII.

is relevant to a given situation. Secondly, the teaching and learning of morality requires experience. Both cases concern the process of mediating between the pure and simple commands of reason and the complex realities of our everyday world. Kant assigns this important task to the faculty of judgement.[17] Regrettably, its practical powers are never fully explained. Yet there may be a deeper reason for this. No matter how specific moral instructions are some element of judgement will always be necessary – moral commands cannot 'go all the way down' to make application automatic or superfluous.[18]

¶ IV 389.36 Pure ethics is naturally of great theoretical interest to moral philosophers,[19] but that is not all. It serves to improve moral practice by clearly revealing and demonstrating the principles of good action.[20] Only a pure moral philosophy can inspire human beings to act morally. This theme is further developed at the beginning of Section II (see IV 410.19).

Kant believes that merely intending what happens to coincide, possibly by chance, with the act demanded by moral laws is morally insufficient. One must consciously intend to perform the right act *because* it is commanded by these laws. If so, we require the clearest possible conception of these laws as well as their authority[21] – which is precisely what a metaphysics of morals is supposed to supply. A 'moral philosophy' that fails to differentiate between the pure and the empirical does not deserve its name. It obscures the correct conception of morality and thus threatens the very possibility of moral action. Kant believed

17 Judgement is defined in the *Critique of Pure Reason* as the capacity to subsume under rules (*casus datae legis*), i.e. to distinguish whether something stands under a given rule or not (A 132/B 171). Much recent work on moral judgement has been inspired by Barbara Herman's *The Practice of Moral Judgment* (Harvard University Press, 1993).

18 See A 133/B 172 and *Theory and Practice*, VIII 275.8–17; cf. also *Critique of Judgement*, V 169.1–14. Moreover, judgement is needed to resolve apparent conflicts of moral commands. This is always possible; see *Metaphysics of Morals*, VI 224.9–26.

19 Note that 'speculation' (IV 389.37) refers to the realm of thought or contemplation, as opposed to 'practice', which concerns action. The word does not have connotations of uncertainty or mere conjecture.

20 The structure of the first sentence of this paragraph is not entirely perspicuous. Kant has not yet mentioned the single 'clue' or 'guiding thread' (*Leitfaden*) or 'supreme norm' to judge all moral principles to which he is apparently referring at IV 390.3 (the categorical imperative). He may, however, be associating the 'source' (IV 389.37) of practical principles with the principle of autonomy, the proper metaphysical formulation of the categorical imperative.

21 It would seem that a metaphysics of morals is not, or at any rate not primarily, concerned with formulating a standard of moral action, which can be done at a more basic philosophical level. The metaphysical account of self-legislation in a realm of moral equals that emerges towards the end of Section II is meant to *inspire* human beings to act morally – something a mixed, impure moral philosophy will fail to do. See Appendix F.

in the principle that 'ought implies can', or rather that moral laws are valid as commands only if the agent is capable of acting accordingly. Ignorance and confusion incapacitate. The morally correct action must at all times be available to human beings if they are to be responsible for their moral failures as well as successes; it must not be dependent on contingent factors. The distinction between action that merely coincides with what the law demands and action that is done for the sake of the law is officially introduced in Section I (IV 397–9).

¶ IV 390.19 Christian Wolff proposes to lay the foundations of practical philosophy in his massive two-volume *Philosophia practica universalis*. By Kant's standards, however, Wolff's work fails to be a metaphysics of morals. It is not confined to the pure part of moral philosophy; nor does it lay bare the a priori sources of moral agency. As the title indicates, Wolff's book is concerned with human volition and action quite generally, and the moral motive of acting solely for the sake of the law is not sufficiently distinguished. The moral motive seems to be one motive amongst many others, not – as in Kantian ethics – a highly peculiar incentive that whenever necessary can and must overcome its inclination-based competition. In Section II of the *Groundwork* Kant purposely proceeds from just such a general definition of the will (IV 412.26–30), and on this basis develops his doctrine of the categorical imperative by separating the will's pure and empirical modes of volition. Kant considers himself to have succeeded at a task Wolff did not even recognise.

3 *The project of* grounding *a metaphysics of morals*

¶ IV 391.16 Kant has so far been arguing for the urgency of developing an unfamiliar, novel kind of pure moral philosophy, a 'metaphysics of morals'. It is only now that he turns to the prior project of laying the foundations of the new discipline. The *Groundwork* is exclusively concerned with the latter task.

It is important to keep the two projects separate.[22] For if the *Groundwork* leads up to, but is largely not itself part of, a metaphysics of morals, it is not subject to the lofty rules of pure philosophy that in the course

22 That is why, reflecting the German use of the preposition, the title of the book should probably be rendered *Groundwork for* – rather than *of – the Metaphysics of Morals* (*Grundlegung zur Metaphysik der Sitten*). This is now becoming more common (see the new translations by Wood, Zweig and Denis).

of the book Kant frequently seems to flout. Rather, the *Groundwork* is bound to be 'impure', at least initially, while the detritus of popular ethical theories is cleared away. The *Groundwork* must separate the pure and the empirical elements of human volition. For this purpose, it first investigates and determines the principles of moral philosophy proper on the basis of 'common rational moral cognition' and 'popular moral philosophy' (see IV 393 and IV 406 respectively). It belongs to a 'Metaphysics of Morals' at most in the sense in which a preface is part of a book.

There are two reasons, Kant reports, why he prefers the more modest title of the present 'Groundwork'[23] to that of a 'Critique of Pure Practical Reason'. First, such a critique is less urgent. Unlike its theoretical counterpart, pure *practical* reason works perfectly well when left to its own devices. It does not generate transcendental illusions, or get entangled in contradictions, which a critique would then have to resolve. Pure practical reason is not, in Kant's sense, 'dialectical'.[24] *Practical* reason runs into trouble only when we investigate pure and empirical volition combined. This is further explained below (IV 404.37–405.35). Secondly, the very project of a second critique raises the problem of the *unity* of practical and theoretical reason, which it would have to demonstrate, represent or portray (*darstellen*, see IV 391.27) under a common principle. This is an arcane project frequently mentioned in Kant's *oeuvre* but never fully executed. It is difficult to say anything useful about this issue in a concise philosophical commentary on the *Groundwork*.[25] However, the present passage at least provides an instructive clue to the nature of the project. 'After all' – Kant continues – 'it can be only one and the same reason, which must merely differ in its application.' That which is to be demonstrated would seem to be not that there is only a single, unitary rational faculty rather than two. This

23 The term is clearly associated with Kant's critical project; see *Critique of Pure Reason* A 3/B 7, where pre-critical metaphysics is said to pay insufficient attention to the foundation (*Grundlegung*) of the discipline.

24 This sentiment still reverberates in the Preface of Kant's second *Critique* of 1788, the *Critique of Practical* [sic] *Reason*; see V 3. However, there is now a 'dialectic of pure [sic!] practical reason'; see V 107. On the face of it, Kant has changed his mind. According to H. Klemme, the discovery of the dialectic of (pure) practical reason decisively influenced Kant's decision in late 1786 or early 1787 to write a second *Critique*; see his introduction to the Meiner edition, p. XIX.

25 For more extensive discussions on the 'Unity of Reason', see Paul Guyer, 'The Unity of Reason: Pure Reason as Practical Reason in Kant's Early Conception of the Transcendental Dialectic', *The Monist* 72 (1989), 139–67; Pauline Kleingeld, 'Kant and the Unity of Theoretical and Practical Reason', *Review of Metaphysics* 52 (1998), 311–39; Susan Neiman, *The Unity of Reason* (Oxford University Press, 1994); and Angelica Nuzzo, *Kant and the Unity of Reason* (Purdue University Press, 2005).

is a presupposition of the project of a critique of reason. Similarly, the second *Critique*'s doctrine of freedom as the 'keystone' (V 3.25–6) that completes and sustains the system of reason depends on this assumption. Practical reason, Kant argues, could not legitimately answer the metaphysical questions left unresolved by theoretical reason if the faculty itself were not unitary, i.e. if, though in a different mode of operation, reason were not answering *its very own* questions (V 121.4–5). Yet it can. (This is the doctrine of the 'primacy of practical reason', see V 119.) By contrast, the project alluded to at present is the *execution* of a quasi-Leibnizian philosophical system in which everything can be deduced from a single highest principle. Of course, such a system would indirectly confirm the existence of a single rational faculty. It would also be rooted in the unity imposed by this faculty upon its objects (see R 5553, XVIII 221–9, and *Critique of Pure Reason* A 302/B 359). It is the unity of the perfect philosophical edifice, that – echoing the 'Dialectic' of the first *Critique* – is declared to be 'the undeniable need of human reason, which finds complete satisfaction only in a complete systematic unity of its cognitions' (V 91.5–7) in the second.

The second sentence of this paragraph provides us with two, possibly three examples of the kind of terminological infelicities we so frequently encounter in Kant's writings. First, Kant calls the *Critique of Pure Reason* the 'Critique of Pure Speculative Reason'. This is not, of course, the title under which the book was published and, more importantly, the first *Critique* was not, at least originally, meant to be confined to preparing a new metaphysics in the sphere of theoretical philosophy.[26] At the time of writing the *Groundwork* Kant apparently considers his earlier discussion of ethical topics an insufficient foundation of a metaphysics of morals.[27] Secondly, according to his earlier classification of philosophical disciplines, Kant should strictly speaking have called the speculative discipline 'metaphysics of nature', not simply 'metaphysics'. Kant's mistake illustrates the fact that these two expressions are commonly used interchangeably. Even the professed founder of the metaphysics of morals apparently needs some time to get used to this novel concept. Thirdly, Kant mentions the project of a 'Critique of *Pure* Practical Reason', which he explicitly rejects on the very first page of the Preface of the *Critique of Practical Reason* (V 3). In the *Groundwork*, Kant needs

26 See especially the section on the Highest Good, A 804–19/B 832–47.

27 By the time of the second *Critique*, the 'Critique of Pure Speculative Reason' and the 'Critique of Practical Reason' appear to be parallel projects; see e.g. V 3.4 and V 16.22, V 42.20–1.

a rudimentary critique of pure practical reason to show how synthetic practical principles are possible; in the second *Critique* he says that no such critique is needed. Part of the philosophical reason for this peculiarity has already been mentioned. It stems from an ambiguity within the very concept of a 'critique' that becomes fully apparent only when one considers the kinds of 'dialectic' threatening the theoretical and practical employment of reason (see IV 405 below).

¶ IV 391.34 All philosophical subtleties are fortunately contained in the present 'preparatory work' (*Vorarbeitung*) for us to be able to enjoy a later book on the metaphysics of morals – presumably an elevating narrative about the autonomy of rational agents in a moral commonwealth – without further irritation. According to the *Metaphysics of Morals* of 1797 the comparative popularity of such a work is due to the fact that in moral – as opposed to speculative – matters sound common reason is naturally on the right track (VI 206.21–8).[28]

¶ IV 392.3 Kant argues that the foundation of a metaphysics of morals is a complete project in its own right, separate from the eventual pure branch of philosophy it seeks to establish. The *Groundwork*'s dual task consists in the identification or discovery of the highest principle of morality (*Aufsuchung*), and in its subsequent corroboration or justification (*Festsetzung*).[29] This happens in Sections I and II and in Section III respectively.

Two things should be noted. First, Kant does not seem to have envisaged that there might be more than one supreme moral principle. This may partly be due to the fact that the prospect of a plurality of 'supreme' ethical principles is philosophically undesirable. Principles of equal normative status might issue conflicting commands, and one would constantly be seeking an 'even more supreme' criterion to mediate between them. Secondly, finding what is implicitly considered a normative principle does not establish its validity. That is why in Section III a 'deduction' of the claims implicit in moral concepts is needed to corroborate the results of Sections I and II.

28 On true and false popularity in moral philosophy see also IV 409.20–410.2 below.

29 The official task of the *Groundwork* is repeated almost verbatim in the Preface of the *Critique of Practical Reason* as 'stating' and 'justifying' a 'determinate formula' of morality (V 8.11), i.e. the categorical imperative or perhaps the formula of autonomy. This would be odd if, as some people think, Kant had come to consider the deduction of the categorical imperative in Section III of the *Groundwork* a complete failure.

Kant adds that he prefers this thorny method of strict philosophical argument to merely demonstrating the implications of his new principle in practical use. Application without prior justification, he fears, might create an impression of partiality and rhetoric. This argument is not entirely convincing.

¶ IV 392.17 The brief last paragraph of the Preface is as controversial as it is essential for our understanding of the book. The first section analyses common moral convictions – the concepts of a good will as the only unconditional good and then the concept of duty (see IV 393.5 and IV 397.7) – and thereby arrives at a first philosophical formulation of the supreme principle of morality (IV 402.7–9). The second section makes a fresh start. It presupposes the foundation of popular moral philosophy – the concept of the will in general (IV 412.28) – and proceeds, again largely by means of philosophical analysis,[30] to introduce the concept of autonomy (self-legislation) of the will, which is central to the project of a metaphysics of morals. This section thus ultimately reveals the source – see IV 405.25 – of the supreme principle: the will itself. The third section steps back to substantiate the results of the previous section. It takes 'freedom' of the will as its starting point (IV 446.5). Kant seeks to defend the categorical imperative – the principle of autonomy, which was the result of the analysis of Section II – as a valid normative principle, a task that leads away from the metaphysics of morals to a 'critique' of pure practical reason. This involves a critical investigation of the 'sources' of the principle of autonomy (IV 392.20, cf. A xii). Foreshadowing the second *Critique*'s doctrine of the 'fact of reason', Kant returns to commonly held beliefs in the course of this.[31] We have already seen that the reality of a concept cannot be demonstrated by means of analytic judgements. The validity of the categorical imperative was merely presupposed in Sections I and II. The justification of morality in Section III must therefore resort to synthetic judgements. The structure of the *Groundwork* thus bears some resemblance to the Analytic of the *Critique of Pure Reason* up until the end of the

30 See IV 445.7–8. 'Largely' because the combination of the first and second variant formulations at IV 431.9–18, which leads to the notion of autonomy, is probably intended to be synthetic; and the change from the concept of the good will to duty marks an important hiatus in the course of Section I. The analytic procedure does not appear to demand a continuous, sustained argument that begins on the first page of a section and ends on the last.

31 The doctrine of the two 'standpoints', IV 450.37, and the commonsense confirmation of the 'deduction', IV 454.20.

Transcendental Deduction. Its 'clue' or 'guiding thread' (see A 66/B 91) are the concepts of good will and duty.[32]

The three sections are – in line with the purpose of a 'groundwork' – *transitions* in the literal sense of the word, not blueprints of the disciplines to which they lead. They presuppose something as given, and lead to the rudiments of, first, moral philosophy (the definition of duty and its principle, IV 400–3), secondly a metaphysics of morals (a theory of autonomy, IV 431–6 and IV 440–4) and, thirdly, a (partial) critique of pure practical reason (the 'deduction' of the moral law, IV 453–5) respectively.

BIBLIOGRAPHY

Bittner, Rüdiger. 'Das Unternehmen einer Grundlegung zur Metaphysik der Sitten', in *Grundlegung zur Metaphysik der Sitten*, ed. O. Höffe (Klostermann, 1993), 13–30

Guyer, Paul. 'The Unity of Reason: Pure Reason as Practical Reason in Kant's Early Conception of the Transcendental Dialectic', *The Monist* 72 (1989), 139–67

Kleingeld, Pauline. 'Kant on the Unity of Theoretical and Practical Reason', *Review of Metaphysics* 52 (1998), 311–39

Klemme, Heiner. 'Einleitung', in *Immanuel Kant. Kritik der praktischen Vernunft* (Felix Meiner, 2003)

Kuehn, Manfred. 'Einleitung', in *Immanuel Kant. Vorlesung zur Moralphilosophie*, ed. Werner Stark (De Gruyter, 2004)

Neiman, Susan. *The Unity of Reason* (Oxford University Press, 1994)

Nuzzo, Angelica. *Kant and the Unity of Reason* (Purdue University Press, 2005)

Schönecker, Dieter. 'Zur Analytizität der *Grundlegung*', *Kant-Studien* 87 (1996), 348–54

Wolff, Christian. *Philosophia practica universalis* (Renger, 1738/9; repr. Olms, 1971/9)

32 According to H. J. Paton, the *Groundwork* follows the analytic and synthetic methods as outlined at *Prolegomena* IV 274–5; see H. J. Paton, *The Categorical Imperative. A Study in Kant's Philosophy* (Hutchinson, 1947), p. 27. This is a natural assumption to make. Kant's all too brief description of the 'research programme' pursued in the *Groundwork* is reminiscent of the 'regressive' and 'progressive' methods as outlined in the *Prolegomena*; and the analytic method of the *Prolegomena* excludes questions of justification, as does Kant's analytic way of proceeding in Sections I and II of the *Groundwork*. However, Kant expressly warns us not to confuse analytic and synthetic method on the one hand with analytic and synthetic judgements on the other (*Prolegomena*, IV 276 fn.): the analytic method often proceeds by means of synthetic judgements – whereas Sections I and II of the *Groundwork* analytically spell out the implications of – admittedly synthetic – propositions. Moreover, Kant explicitly conceives of Section III as justifying the categorical imperative as a synthetic proposition a priori (see e.g. IV 419 and IV 444–5); and the project of a (rudimentary) 'deduction' of a synthetic principle does not proceed by means of the 'synthetic method'. These points make it seem unlikely that Kant explicitly employs the method of presentation set out in the *Prolegomena*. Cf. D. Schönecker and A. Wood, *Kants 'Grundlegung zur Metaphysik der Sitten'* (Schöningh, 2002), p. 14.

Section I: Transition from common to philosophic moral cognition of reason

Kant presupposes what he takes to be the central belief about the nature of value that, at least implicitly, is universally acknowledged to be true: only a good will is good absolutely and without restriction. However, an analysis of the nature of goodness does not advance the project of grounding a new moral philosophy. Any substantive conception of value leads to what Kant later calls 'heteronomy': causal regularities that help us to realise a pre-given end, not to a formal law that first defines moral value. That is why he switches to analysing the cognate concept of duty. The discussion of clear cases of action 'from duty' – the suicidal person who nevertheless preserves his life, the beneficent act of someone not naturally sympathetic, the gout sufferer who still takes care of his long-term well-being – reveals that duty is characterised by a certain robustness in the face of opposing inclination. Actions have moral value only if they are done from duty; and duty is defined as the necessity of an action from reverence for 'the law'.

What kind of law can this be? Kant claims to have shown that it must be purely formal, because the matter of willing – its intention or purpose – has been shown to be morally immaterial. He proceeds to give a first formulation of the supreme principle of all moral action, illustrated only briefly by means of the example of a lying promise and not yet christened with its official name, the 'categorical imperative'. Reflections on the nature and purpose of the transition to moral philosophy just accomplished conclude the section. As the level of philosophical sophistication is reached only gradually in the course of Section I, it is much less technical in vocabulary and style of exposition than the more professional second section.

1 On the unconditional value of a good will

¶ IV 393.5 The celebrated opening sentence of the *Groundwork* represents a first bold answer to the guiding question of pre-Kantian, especially ancient ethics: What is the highest good?[1] Kant replies: not

1 For this precise expression, applied to the good will, see IV 396.25 below. The subject matter of moral philosophy is defined by moral goodness, even if it is ultimately to be defined in terms of a formal law. The study of moral philosophy must start with the notions of what is good or bad in itself; see explicitly R 7216, XIX 287–8, dated 1780–9.

talents, qualities of temperament, or gifts of fortune;[2] not happiness, as philosophers have assumed since classical antiquity; but rather *a good will*.

This answer accords with what he considers to be common ethical knowledge. In a series of lectures on moral philosophy that he gave in the winter semester of 1784–5, when the *Groundwork* was in the press, Kant claims that 'everyone knows that nothing in the world is entirely good without restriction save a good will' (*Mrongovius* II, XXIX 607.15–16).[3] Even so, the opening sentence should not be thought simply to rest on the empirical findings of, for example, an opinion poll. It is obviously normative (see IV 406–8). According to Kant, ordinary moral practice merely implicitly commits all mature human beings to valuing the good will above all other goods, and on reflection everyone will appreciate this truth. At IV 397.3, he declares this conception of a good will to 'reside' in natural, sound understanding. It does not need to be 'taught' but merely 'clarified' (*aufgeklärt*).[4] Common practical reason rejects the opinions of those who have come to esteem things other than a good will unconditionally and without restriction. Returning to the theme of common moral consciousness in Section III, Kant intimates that everyone, even the most perverted villain, would on reflection agree with the opening statement when confronted with shining examples of virtuous conduct (IV 454.21–2).

Kant will shortly abandon the concept of moral goodness in favour of the more informative concept of *duty* as the basis of his analysis in Section I (see the hiatus at IV 397.1–10). Morally good action is not action for the sake of some perceived value. However, the notion of a good will as rooted in common moral thought is central to the *Groundwork*. Action that proceeds from a sense of duty is the paradigm case of good human willing; and Kant returns to the theme of the good will time and again throughout the book.[5] His parenthetical remark

2 *Glücksgaben*, IV 393.13–14. These are external goods we obtain through fortune (*Glück*), as opposed to internal goods like talents and qualities of character as 'gifts of nature' (*Naturgaben*, IV 393.12). The word does not refer to *Glückseligkeit* (happiness).

3 The classical origins of the opening statement are made explicit in the same set of notes: 'The ancient Greeks concentrated the determination of the principle of morality on the question: What is the highest good? Among all that we call good, the major portion is good in a conditional sense, and nothing is good without restriction, save the good will' (XXIX 599.23–6). See also Kant's critique of ancient accounts of the *summum bonum* in the second *Critique*, V 64.25–34, and his description of happiness as a popular candidate for the status of the highest good in the third, V 208.22–5.

4 'Elucidated', 'brought to light' – which shows the *Groundwork* to be a work of the Enlightenment (*Aufklärung*).

5 For instance, a statement of the 'principle' of the unrestrictedly good will concludes the analysis of Section I (IV 402.3–4, IV 403.18–33); cf. IV 397.2, IV 426.10 and particularly the

that even 'beyond the world' there can be no unconditional good other than the good will alludes to the fact that strictly speaking a good will – like the moral quality of human action in general – is not part of the world that we experience.

What exactly is a good will? The first paragraph indicates that Kant is thinking not of isolated good intentions or individual attitudes but rather of good moral volition overall:[6] a morally good character (which will then express itself in good acts of will, and consequently good action). The basic idea of the opening passage then seems to be the following. If one were to ask sound human reason what ought to be considered the incomparably best thing it is at all capable of conceiving, the reply would be: it is a being whose volition is morally good *all round*. Yet common moral consciousness is only dimly aware of the corollaries and implications of this thesis. That is why, by philosophical standards, both the concept of the will and the concept of goodness at first remain comparatively vague. They are clearly defined only in the further course of the *Groundwork*.

Kant knows that the notion of a 'highest good' is ambiguous. Accordingly, his full reply to the fundamental question of classical ethics is rather more complicated than the opening sentence suggests. The good will is indeed the supreme and only unconditional good. However, according to a view developed in the *Critique of Practical Reason*, the 'highest good' – the perfect, complete, comprehensive good for a person to which nothing of value can be added to improve it – is the union of virtue and happiness (V 110.33). It should be noted that, *in nuce*, the idea of a (comprehensive) highest good is in place long before the second *Critique*.[7] If Kant does not dwell on it in the *Groundwork*, it is because, as he puts it in the essay on *Theory and Practice*, as far as 'the question of the principle of morals' is concerned, the doctrine of the

summary of the analytic part of the *Groundwork* at the end of Section II, IV 437.5–20; see also IV 443.27 and IV 444.28–34. Kant returns to the theme of good volition in Section III, IV 447.13; in the course of which we also learn that the good will does not belong to the world of experience and sense, but rather to a special realm that is governed by the laws of reason (IV 455.4). From this perspective, the *Groundwork* is a treatise on the value of action, rather than an exercise in 'deontology'. However, the matter is complicated by the fact that moral action is not done for the sake of its goodness. Rather, it is action for the sake of duty, or its law, which first constitutes goodness; see the second *Critique*'s chapter on the 'object' of pure practical reason, especially V 62.36–63.4. This is one reason why Kant temporarily abandons the notion of the good will at IV 397.1–10 below.

6 For a discussion of these three possibilities see K. Ameriks, 'Kant on the Good Will', in *Interpreting Kant's Critiques* (Clarendon Press, 2003).

7 See e.g. the lectures on moral philosophy (*Collins*, XXVII 247, and *Mrongovius* II, XXIX 600) and *Critique of Pure Reason* A 814/B 842.

highest good 'can be completely passed over and set aside (as episodic)' (VIII 280.5–8). It is irrelevant to the task at hand. Yet already in the *Groundwork* Kant expressly denies that the good will is the sole and complete good (IV 396.24–6); and there is a hint of the comprehensive notion of the highest good in Kant's ideal of a 'kingdom of ends' introduced towards the end of Section II (IV 433.16). Other things can be good, all things considered – if not *absolutely* good, i.e. unconditionally good in all respects or circumstances. The good will is not even the only good that is non-instrumentally good: happiness, the favourite highest good of ancient philosophy, is similarly pursued just for its own sake. It is precisely this final nature that makes happiness the main rival of (categorical Kantian) morality for the position of the supreme good at the foundation of the highest principle of human conduct in ethical theory. Yet the good will is the only good that can reasonably be considered good *without qualification*. All other candidates – like talents, social or personal goods, living in lucky circumstances, even happiness – are good only conditionally. Some of them are merely good instrumentally, as a means towards other goods; and the *objective* value of any of these factors depends on whether they are attached to a good will or not.[8]

¶ IV 393.25 Ancient virtues like moderation (*Besonnenheit*, σωφροσύνη), self-control (*Selbstbeherrschung*, presumably ἀνδρεία) and calm reflection (*nüchterne Überlegung*, σοφία) are 'conducive' to the good will but its goodness does not depend on them. They are rational and 'appear', as Kant cautiously puts it, to be part of the 'inner' – not 'moral' – worth of a person,[9] but unlike the good will they are not good under all circumstances. They can, for instance, be recruited by a bad cause. Note that Kant's demotion of these qualities depends on his stripping them of the moral significance that for the ancients they naturally possess. (This cannot easily be done with the fourth Platonic virtue of justice, accordingly excluded from Kant's list.) The widely held doctrine of the 'unity of the virtues' even entails that a virtue, almost by definition, cannot be misused. As so often, Kant is relying on the connotations of the German terms, rather than careful exegesis of the sources.

8 Cf. R 6890 (XIX 194–5) and Christine Korsgaard's 'Two Distinctions in Goodness', in *Creating the Kingdom of Ends* (Cambridge University Press, 1996).

9 Kant does not possess a notion of 'moral value' other than that of the moral worth of good volition (*pace* T. Hill, 'Editorial Material', in Zweig's translation of the *Groundwork*, pp. 29, 34 and 267). There is only one word in the original German text. Something possesses moral 'worth' or 'value' when it is morally good. I shall interchangeably use both expressions as translations of *moralischer Wert*.

Why does Kant think a cold-blooded villain more detestable than a more passionate criminal, even if the intentions of both are equally bad? Presumably because he considers cold-blooded deliberation a misuse of the rational faculties that – like any other moral agent – a villain possesses.[10] He is not just swept along by the tide of passion, however willingly. Rather, our villain acts deliberately and with great care, wilfully defying the moral law. Moreover, his cold-bloodedness makes it easier for him to act from duty. He faces no strong inclinations that would have to be overcome. By contrast, someone who is not by nature sober and moderate may have to struggle hard to become a morally good person.[11]

¶ IV 394.13 Kant now introduces an idea that is to dominate the discussion of duty in the course of Section I: in concrete cases of conflict the goodness of a morally good will does not just *outweigh*, it even *silences* any other possible varieties of practical value.[12] Compared to the necessity of the command of duty – see IV 400.18–19 – nothing that stands in its way can have any normative weight at all. The reason is that morality commands actions directly, without presupposing any end, purpose or intention – that which is to be effected by an action – as all other kinds of action do. When a moral command applies, these pre-given purposes are rendered irrelevant. They will be 'readmitted' if they turn out not to conflict with the moral command. Otherwise, a state of affairs can at best be good for the agent, i.e. part of his or her happiness, but may still be bad overall. The thesis of the unique unconditional goodness of good willing is therefore much stronger than modern theories of moral 'overridingness'.

A morally good will is characterised by volition alone. If it was that which is to be realised that marks out a good will, it would not be good unconditionally, but instrumentally, as a means towards a certain end.[13]

10 This idea seems to lie behind the asymmetry noted by Wood, *Kant's Ethical Thought*, pp. 24–5. It is not simply the addition of an otherwise good thing that makes the calculating scoundrel immediately more abominable, but his misuse of reason. This is confirmed at IV 454.21–2, where we encounter this character again: Kant characteristically adds the caveat that the villain wishes to be moral when faced with examples *if only he is otherwise accustomed to the use of reason*.

11 In the lectures on moral philosophy, Kant similarly praises doing good in a calm and orderly fashion. He warns his students not to confuse this with 'frigidity' (*Kaltsinnigkeit*), the want not of a mere emotion but of practical love; see *Collins*, XXVII 420.27–30.

12 On the 'silencing' effect of morality, see *Metaphysics of Morals*, VI 481.34–6.

13 Kant dismisses what we call 'objective' consequentialism – the idea that the value of an action depends on its de facto effects. He also rejects 'subjective' consequentialism – the thesis that the value of an action depends on its intended or foreseen consequences – explicitly so when

However, this does not mean that means and effects are unimportant. A good will is likely to produce agreeable results.

In parentheses, Kant warns us not to confuse the good will with a mere ineffective wish. For a will to be good it must summon the requisite means.[14] That is why in the lectures on moral philosophy 'perfection', though rejected as a moral principle, is still said to be 'indirectly' moral.[15] A good will 'needs the completeness and capacity of all powers to carry out everything willed by the will' (*Collins*, XXVII 266). A morally good will must be careful and circumspect. The way to hell may be paved with individual 'good intentions', but these are unlikely to be the intentions of a will that is good all round, i.e. the individual intentions of a person with a good character.

This thesis seems plausible even on the background of commonly held moral views. No one can reasonably call morally defective the behaviour of those who summon all means and without any fault of their own fail to bring about the intended end. That which is distinctive of *moral* action – the worth of their willing – is not affected by their lack of success. Kant explicitly opposes a prominent variant of what in our contemporary debates is known as 'moral luck'. A good will is likely to be useful; but it is not good *because* it is useful. Its value would not be affected by an accidental lack of utility. The usefulness of the good will can at best help us to persuade those who wish to ground moral philosophy in utility, rather than the dignity of the good will.[16]

Neither the realisation of the end of a single inclination, nor the satisfaction of the 'sum of all inclinations' (IV 394.17–18, Kant's definition of happiness), is good absolutely. The idea is roughly the following. We are happy – we *feel* happy – when we get everything we desire. Kant's formal notion of happiness is quite different from much more substantial ancient notions of *eudaimonia*, and he is hardly, if ever, aware of the

he returns to the topic at IV 399–400 below (the 'second proposition'). It is not the intended ends or the purposes pursued with an action that define a morally good will, but rather the *grounds* that make us elect to pursue them. In the case of morally good volition, what is at stake is the willed action itself, not the action's 'object' or its realisation (see IV 413 fn.).

14 Morality selects, rather than presupposes, ends, but in either case an agent committed to an end must rationally will the means required to realise it. He must resort to 'technical' hypothetical imperatives (see IV 415, IV 417 and Appendix D). Note that the distinction between the moral and the non-moral is upheld. The action, described as taking a certain means such as writing a cheque, is conditionally good; the action described in moral terms as helping a friend is unconditionally good.

15 As at IV 399.3 below, the word 'indirect' indicates the choice of means.

16 The example of a will that 'by a special disfavour of fortune or the niggardly provision of a stepmotherly nature' lacks the resources to achieve anything (IV 394.19–20) points to the controversial 'strategy of isolation' employed later in Section I (IV 397–9). Taking away what is inessential to good volition brings to the fore that which is special about it.

difference. His attacks on the 'eudaemonism' of ancient ethical theories, like his demotion of the three cardinal virtues in the previous paragraph, is therefore rather quick. Kant's conception of happiness is problematic also in that we often feel particularly happy when something good happens to us unexpectedly, without our having pursued a certain purpose and without an obvious inclination that has been satisfied.

Up to this point, Kant's analysis of the good will has been highly abstract. We do not yet know what characterises a good will; nor whether a good will as we conceive of it does in fact exist. These questions will be considered in the further course of the *Groundwork*. Yet some important preliminary decisions have already been taken. Kant has argued for the primacy of good moral willing over all other goods. The moral worth of an action cannot depend on its object, not on that which one wants to bring about, but rather on the reason why and the way in which one wills it – despite the fact that the precise quality of the good will is as yet undetermined. Thus Kant has already established one important result. An ethics of the good will cannot be grounded in the effects or consequences of an agent's willing; and all ethical systems that make the effects of actions on personal or universal happiness their highest principle appear erroneous right from the start. Moreover, the special dignity of moral goodness points to the fact that a will cannot be more or less morally good; it is either good absolutely or not good at all.

2 *A morally good will, not happiness, is the natural purpose of reason*

¶ IV 394.32 The thesis that only a good will is absolutely good is a foundational part of the 'common moral knowledge' referred to in the heading of this section. Yet it also raises suspicions in certain quarters because it seems to detract from the obvious utility of what a will can effect. Kant now tries to defend his thesis that moral goodness is the primary purpose of reason, and to this effect employs an argument that turns on the alleged purpose of nature as 'assigning reason' to the human will 'as its sovereign' (IV 395.1).[17]

Kant's preliminary defence of the authority of morality as rationality does not affect the question – bracketed in Sections I and II – whether human beings possess pure practical reason and thus are capable of

17 See also parallel passage in the *Mrongovius* II lectures (XXIX 640.14–22) and the *Critique of Practical Reason* (V 61.32–62.1) and the slightly more favourable treatment of reason-guided happiness in the 1784 *Idea for a Universal History from a Cosmopolitan Point of View*, VIII 19–20.

morally good action at all. He is, for the moment, presupposing that we are capable of both pure and empirical practical reason and argues for the superiority of the former. These arguments proceed at the level of conventional wisdom, which is enamoured with teleological reasoning. They are amongst the weakest that can *actually* be found in the body of Kant's works.[18] The present four paragraphs form an excursus. Kant resumes the analytic project of Section I of the *Groundwork* afterwards, at IV 397.1.

¶ IV 395.4 Kant's argument relies on the principle of traditional teleology that nature does nothing in vain or without a purpose.[19] If happiness – not morally good willing – were the greatest and unrestricted good, practical reason would be superfluous because in principle instinct would be a more suitable instrument for making us happy. (A creature guided by natural instinct might still be endowed with reason, but this would have to be an exclusively theoretical rational faculty that has no bearing on its actions, not even on the choice of means to ends determined by nature.) Yet nature, Kant argues, has given us practical reason (a will); and this must have happened for a purpose. If not to make us happy, what was nature's intention when she bestowed on us practical reason? She intended us to perform morally good actions. QED.

There is more than one thing that is questionable about this argument. First, Kant grounds his conclusion that moral imperatives are justified as commands of reason on the assumption of a wise government of the world and the general purposefulness of nature, which must appear problematic within the general framework of Kant's critical philosophy. Such assumptions can, perhaps, be justified on the basis of moral conviction,[20] but they have no place in a defence of the normative force of morality as commanded by reason against the claims of eudaemonism. Secondly, Kant's arguments are hardly convincing even if we decide to ignore this difficulty. For a start, he simply assumes that there are only two candidates for the final purpose of practical reason: happiness and morality. If it is not the former, it must be the latter. Furthermore, it is surprising that he should consider a nature that does nothing in vain to have bestowed on us organs that are absolutely

18 W. D. Ross's scathing criticisms are apt for once; see *Kant's Ethical Theory* (Clarendon Press, 1954), p. 13.
19 See e.g. Aristotle's *De partibus animalium* II, 658a8–9: οὐδὲν ποιεῖ μάτην ἡ φύσις.
20 See *Critique of Judgement*, V 442–5.

perfect for their intended purpose. Other natural instruments we possess, for example our eyes, are good but by no means flawless. If so, why should we expect practical reason to be without fault if its primary purpose were that of making us happy? Moreover, it can hardly be taken for granted that reason must be a worse instrument for the purposes of happiness than instinct. Does not reason assist us in understanding and solving many practical problems that would defy a creature guided solely by instinct? (A mouse would curse, if curse it could, its animal instinct when caught in a trap. Reason is needed to understand its mechanism; and even cheese does not always make a living creature happy.) Conversely, it is obvious that our practical reason cannot fulfil the high standard of perfection in the moral realm. Human practical reason functions even less perfectly in the field of moral action than – allegedly – instincts do in serving to satisfy inclinations. For even if we share Kant's optimism regarding the human powers of moral cognition, there still remains a wide gap between judgement and volition. It is one thing to judge an action right; it is quite another thing to act accordingly.

The only good excuse for Kant's teleological excursus lies in the preliminary nature of the present argument. Its purpose is to deflect teleological criticisms of the special status of a morally good will. Kant therefore assumes a background that must otherwise seem problematic. The passage is rooted in 'common moral cognition of reason', rather than rigorous philosophy, and is later to be replaced, within the broader justificatory project of Section III, with a more philosophical – if hardly less problematic – account of the authority of pure practical reason: an account that relies on the idea that we are members of two worlds and, in cases of conflict, must side with the superior intellectual world (see IV 453.32).

¶ IV 395.28 The subsequent thought is similarly unconvincing. Do those who consciously try to lead a happy life soon start to hate reason, as Kant suggests? Do they really envy 'the more common run of people', who are still largely guided by instinct in felicific matters? Does this point to the 'worthier purpose' (IV 396.10) of one's rational faculties of limiting the pursuit of happiness to moral conditions? Does not 'misology'[21] (the hatred of reason) rather result from an overly

21 A term coined by Plato; see *Phaedo* 89d ff. Kant was familiar with the *Phaedo* through Mendelssohn's adaptation of 1767, but he hardly needed to have read the book to be familiar with the term. Cf. R 2570, XVI 424, on the origins of misanthropy and misology.

zealous pursuit of happiness, which is unlikely to be endorsed by rational reflection? In what we may call his 'genealogy of freedom', Kant explains how incipient practical reason leads to new desires (*Conjectural Beginning*, VIII 111–12). Even so, it would be odd if the proliferation of unnecessary desires persisted once mature practical reason has learnt to accord priority to its moral task.

¶ IV 396.14 For Kant, the 'true vocation' (IV 396.20) of our practical rational capacities thus consists in being good immediately, i.e. morally, and not just instrumentally, for the sake of something else. Without (pure) practical reason, unrestricted moral volition would be unthinkable because all action would be directed at the satisfaction of some natural need.

Kant calls a will that is good in itself – and later, by analogy, God (IV 409.1) – the 'highest good' (IV 396.25) in the sense of the *supreme* good. The 'highest good' in the sense of the *comprehensive* or *perfect* good is the coincidence of moral worth and happiness. If so, the good will is obviously neither the only nor the complete good, even if it is the condition restricting the goodness of everything else, even our desire to be happy. This leads to yet another problem for Kantian teleology. If in nature everything is created with a view to perfection, any imbalance between our need to be happy and moral reason is hardly intelligible. Even if the fulfilment of the true purpose of practical reason brings with it a certain contentment (which of course cannot be an appropriate reason for moral action), this hardly saves the thesis that nature is a perfect teleological system.

One thing, however, is clear. Instinct can no longer work smoothly once reason takes over as our primary guide in the sphere of action. Actions aiming at our own greatest happiness are often rejected by reason, and we must then content ourselves with the second best option. Crucially, identifying this option is a task also assigned to our rational faculties. For Kant, empirical practical reason – later so called – is therefore no more than a cumbersome by-product of our rationality, a poor substitute that tries to take over the role of instinct when it no longer infallibly serves to make us happy. Also, the fact that non-moral ends must now be subjected to the standards of morality does not entail that reason has a completely free choice in adopting such ends. Reason can reject ends on moral grounds and then take things further on that basis; but they must still first be proposed by nature.

3 Elucidation of the concept of duty by means of three propositions

a. The general concept of a good will must be made more determinate by analysing the concept of duty

¶ IV 397.1 We have now been presented with several reasons why a morally good character deserves the limitless estimation commonly conferred on it by sound practical reason, but we are still in the dark about its defining characteristic. What distinguishes a will that is good in the required sense? The answer is to be revealed by an analysis of what it means to act *from duty*.

The concept of duty 'contains' the concept of the good will because in the case of finite beings like ourselves action from duty is the paradigmatic case of good volition. Doing the right thing is at times difficult for creatures whose manifold needs and desires are potentially at odds with morality.[22] Human volition does not as such coincide with what is morally good: we do not always effortlessly do the moral thing of our own accord. In moral action, we do not act for the sake of some perceived good at all – moral action first constitutes the good. That is why we encounter the objective laws of morality subjectively as 'imperatives', as necessitating *commands* of duty (see IV 412.35–413.11). Moral action is action for the sake of duty. It is Kant's hope that the contrast between human volition and moral commands will expose the principle of the good will.

This change of concepts marks a hiatus in the argument of Sections I and II.[23] For a perfect will the laws of morality hold 'analytically'. It contains the motive of good willing (and as it is unaffected by sensuous inclination, presumably nothing else). A will of this type effortlessly and without any threat of internal conflict wills that which is good. By contrast, a will like the human will faces morality as a command with which it does not necessarily comply – despite the fact that of course it ought to. Laws of duty command something not previously contained in the will as desired or good: such commands are 'synthetic'. It is at this point in the argument of Section I that Kant starts analysing moral

22 This is the significance of the somewhat misleading parenthetical comment about 'subjective restrictions and obstacles' (IV 397.7–8): they affect a good human will, not its concept. See Schönecker, *Kant: Grundlegung III*, pp. 34–5.

23 This explains why at IV 447.10–14 he denies that the categorical imperative can be analytically derived from the concept of a good will.

laws as synthetic laws of duty. In fact, Kant's rejection of all theories that take a specific conception of the good as their starting point in the second *Critique* suggests that the moral law cannot be teased out of the notion of good volition at all, presumably because it points in the direction of ends to be realised, rather than formal laws (V 63.11–64.5, cf. V 8.25–9.3).[24] Presupposing any conception of value makes all commands instrumental and moral theory heteronomous. Moral value is first created by acting for the sake of the law.

Theoretical and practical synthetic principles can both be subjected to conceptual analysis. In both cases the result, if more determinate, will still be synthetic. As such, the normative status of synthetic a priori moral commands (their 'possibility') is problematic. This difficulty will be addressed in the synthetic third section of the *Groundwork*.

b. Proposition 1: An action that coincides with duty has moral worth if and only if its maxim produces it by necessity, even without or contrary to inclination

¶ IV 397.11 The following parts of Section I are, to say the least, controversial. Readers have found Kant's pronouncements on the 'moral worth' of actions objectionable; the correct interpretation of the suggested moral psychology is disputed; and there is the riddle of the missing 'first proposition', which has puzzled readers and delighted Kant scholars for generations.

Kant explicitly states only a second and a third proposition (see IV 399.35–400.3 and IV 400.18–19 respectively). Which thesis does Kant intend as the first? A natural candidate would seem to be the idea that only action from duty has moral worth, but this creates huge problems for the interpretation of Kant's suggestion at IV 400.18–19 that the third proposition is a 'consequence' of the two preceding ones.[25] There is, fortunately, another thesis that encapsulates the gist of

24 The chapter on the 'object' of pure practical reason is Kant's reply to a point raised by H. A. Pistorius in his review of the *Groundwork*. See R. Bittner and K. Cramer, *Materialien zu Kants 'Kritik der praktischen Vernunft'* (Suhrkamp, 1975), pp. 144–60.

25 Most interpreters, from Paton (*The Moral Law*, p. 19) to Hill ('Editorial Material', p. 267), identify the first proposition with the thesis that an action is morally good (if and) only if it is done from duty. Exceptions are A. R. C. Duncan, who takes the 'first proposition' to be the opening sentence of Section I (*Practical Reason and Morality. A Study of Immanuel Kant's 'Foundations for the Metaphysics of Morals'* (Thomas Nelson, 1957), p. 59), and D. Schönecker ('What is the "First Proposition" Regarding Duty in Kant's *Grundlegung*?' in *Kant und die Berliner Aufklärung*, ed. V. Gerhardt, R.-P. Horstmann and R. Schumacher (De Gruyter, 2001), vol. III, pp. 89–95); see also his joint commentary with Allen Wood (*Kant's 'Grundlegung'*,

these examples: what one might call the 'principle of non-contingence', which is implied by Kant at the end of each of the subsequent paragraphs. According to this principle, a 'dutiful'[26] action has moral worth if and only if its maxim possesses the property of making it necessary, i.e. generating the action non-accidentally, independently of the agent's current state of inclination or changing external circumstances. As we shall soon see, the 'principle of non-contingence' is the superior candidate because it can account for Kant's contention that the third proposition combines elements of the other two.

It is important to note that Kant's analysis of the now central concept of duty contains an implicit 'rigoristic' assumption. As we learn in the Preface, it is not sufficient, morally, when an action merely happens to coincide with what the moral law commands; it 'must also be done for the sake of the law', otherwise the coincidence with morality is only very contingent and precarious (IV 390.4–6). Kant also assumes that all actions follow some law or other.[27] This explains why he now sets out to find clear cases of action motivated by duty. One cannot hope to gain insight into the nature of morality from an examination of actions from inclination that merely accidentally – we do not quite know why – concur with moral principles but are in fact governed by some entirely different law, for example a prudential guideline. Of course, no such insight is to be gained from an examination of actions that violate duty either. Human beings do not violate their duty for the sake of it – that would be diabolical – but rather to satisfy some particular inclination that happens to conflict with duty (see *Religion* VI 35.20–5). There is no inherent connection between inclination and duty. They are different in kind. Whether an action done from inclination accords with duty or not is therefore ultimately a matter of chance; which is precisely why morality must be put in charge, and inclination pursued only within the limits of the moral. A moral action must be done *because* the law

p. 60), who considers the first proposition to be the thesis that an action from duty is an action from reverence for the moral law. J. Freudiger similarly emphasises the motivational significance of a first proposition but postpones the introduction of reverence until the third (*Kant's Begründung der praktischen Philosophie* (Paul Haupt, 1993), pp. 78–80). The only other possible candidate might seem to be the thesis presented at the beginning of the discussion of duty at IV 397.6–10 that the concept of duty contains the concept of the good will under certain subjective restrictions and obstacles. If so, it is not exactly obvious how the third proposition is supposed to follow from the other two.

26 An action that accords or coincides with duty (*pflichtgemäß*).

27 See the official definitions of the will (IV 412.26–8, IV 427.19–20) and the explicit argument for regularity by virtue of the will's causal nature at IV 446.15–21. Presumably even actions defying imperatives of prudence and morality obey some sort of regularity, if only a technical rule that the end of immediate inclination can be realised thus-and-so.

commands it. Only a clear example of such action *from duty* will advance the present project of identifying its principle.[28]

So, what is an action *from duty*? Consider a situation in which a certain course of action is morally required. Any action will then either coincide with duty or be contrary to duty. There are no permissible actions that coincide with duty in *that* sense – if an action is obligatory it is a fortiori the only permissible action. Kant proceeds to exclude actions contrary to duty, no matter how useful, for the reason already mentioned: their investigation does nothing to advance the search for the supreme principle of morality.[29] Within the class of actions that accord with duty, Kant then distinguishes the following three types (IV 397.14–21):

A. actions (merely) coinciding with duty that are performed as a *means* to satisfying a higher-order inclination directed at some other object;
B. actions (merely) coinciding with duty, performed from an *immediate* inclination towards some object;
C. actions that (coincide with duty and) are performed solely *for the sake of duty*.

Note that there can be action performed *with* concurrent inclination and yet *from* the motive of duty – for instance if someone possesses firm moral principles and also naturally enjoys doing good deeds. In this case – as in the case of a person who has to overcome strong inclination to act from duty – inclination is rendered ineffective by the agent's moral maxim.[30]

Thus because of the different nature of moral and non-moral motivation, a trace of conflict remains even if in effect duty and inclination concur and an action is in this sense 'overdetermined'. Duty commands the willing of an *action* (C) whereas inclination, either directly or indirectly, aims at the *realisation* of a specific objective in the world (B or A).[31] We might say that all inclination is by nature consequentialist.

28 In the *Critique of Practical Reason*, the distinction between actions that merely coincide with duty and those done for the sake of duty is re-cast in terms of 'legality' and 'morality'; see V 71.28–34 and V 81.10–19, as well as V 72 fn., where the 'letter' is distinguished from the 'spirit' of the law.

29 Similarly, Kant does not seem to think that the scrutiny of merely permissible actions can reveal to us the principle of morality. This category of action is not even mentioned.

30 This is one of the points overlooked in Friedrich Schiller's mock criticism of Kant. See Appendix A.

31 See IV 413–14 fn. on different types of interest. The aim of moral actions is contained in the action itself: it is exclusively directed at the act of volition, which then expresses itself in some

Kant therefore maintains that agents must act from duty, even in such cases of lucky coincidence, if their action is to possess true moral worth. Significantly, if an action on moral principle is graced with success, a parallel inclination towards a good effect will still be satisfied, despite the fact that inclination did not influence the course of action taken. An action from duty need not be contrary to inclination, and the mere presence of an inclination coinciding with duty does not eradicate the agent's moral worth if only his maxim has 'moral content'.[32] By contrast, an action that merely coincides with duty – an action of type A or even type B, motivated by an interest in some result and not in moral volition – can never satisfy the command to act dutifully for the right reason, i.e. to perform an action of type C.

There is therefore no need for duty – action for the sake of the law – to be relegated to a mere 'backup motive' that takes over when inclination fails.[33] At IV 400.25–31, we learn that inclinations must not determine what ought to be done. If so, why should they be re-admitted to moral practice when it comes to executing the conclusions of one's reasoning? Why would someone choose to be motivated by inclination *on the grounds* that it happens to coincide with morality? Why not just act on moral grounds? To (mis)use a familiar phrase, an agent who acts on inclination on condition that it coincides with duty entertains *one thought too many*.

At this stage of the argument of Section I, Kant thinks that actions done from immediate inclination (B) are interesting in a way in which actions to which agents are 'impelled' through some other inclination (A) are not. In the latter case, it is said to be 'easy to distinguish' whether an action coinciding with duty was done 'from duty' or 'from a self-seeking purpose'; and Kant continues to say that 'it is much more difficult to note this difference' – *sc*. the difference between morally valuable action 'from duty' and generally action that lacks moral worth[34] – when

external act. It is the mystery of morality how it is possible for us to be directly interested in an action without being interested in its effects. By contrast, non-moral, inclination-based interest is directed not at the action but at the 'object' to be realised, in so far as it is 'agreeable' (*angenehm*) to the agent. See IV 459–60 fn. and second *Critique*, V 20.22. The same distinction is implied at IV 400.19–21, where the effect of an action is said to be the proper object of inclination but not of the moral incentive of reverence.

32 'Moral content' and 'moral worth': it is the moral *content* of the maxim on which we act that makes the action morally *good*. Moral content of a maxim is equated with the commitment to do the morally correct action from duty; see IV 398.19–20.

33 The *locus classicus* is Hume's *Treatise*, Book III, Pt. II, Section I; a similar picture is favoured by virtue theorists like M. Stocker. See his 'The Schizophrenia of Modern Ethical Theories', *Journal of Philosophy* 73 (1976), 453–66.

34 Kant's expression lacks precision. 'This distinction' (IV 397.29) cannot be the specific difference between selfish action and action from duty – the subsequent examples of action

an action coincides with duty and the subject has, besides, 'an *immediate* inclination' (IV 397.20–1), i.e. action of type B. This is offered as a reason why a discussion of A-type action – just like a discussion of action that is contrary to duty – would be unfruitful and is therefore to be 'set aside' at the end of the present paragraph. Kant is saying that a decent act done merely for higher-order selfish reasons *plainly* lacks moral worth, and that agents who act in this manner *obviously* lack a morally good will. By contrast, the idea that action from more immediate motives such as sympathy lack moral worth is much less obviously an expression of common moral consciousness.[35] Defending the value of acting from duty will be much harder. This is the task of the examples contained in the three subsequent paragraphs.

First, however, the shopkeeper (IV 397.21). The case of the prudent salesman who does not overcharge his inexperienced customers concerns the strict duty of honesty towards other human beings.[36] It belongs in category A, soon to be 'set aside'; not in category B, as at first the text seems to suggest. Kant plausibly argues that a shopkeeper is unlikely to treat his customers honestly because he has an immediate liking for them – in which case he might give his goods away for free. The examples discussed at length in the three paragraphs that follow – preserving one's life, beneficent action and caring for one's happiness – are all common cases of type B.[37]

The details of the shopkeeper case are puzzling. It is odd that Kant so carelessly attributes the shopkeeper's behaviour to 'purposes of self-interest' rather than moral conviction. The only apparent reason for this diagnosis is his possessing self-interested grounds of prudence that

commonly done from immediate inclination do not concern people's selfishness, as Wood points out (*Kant's Ethical Thought*, p. 30). It is more likely that the statement should refer to individual acts done from duty or from inclination – but, then, Kant is committed to denying in principle that the moral attitudes of individuals can be discerned. The prudent merchant, for instance, might flatter himself that he is a moral man because after all he never treats anyone dishonestly. It seems more likely that the distinction in question is that of the *kind* of action that possesses moral worth, and that where the distinction is first stated the moralising expression 'from a self-seeking [or perhaps: selfish] purpose' (*in selbstsüchtiger Absicht*) is just a placeholder for a type of action that clearly lacks moral worth. (Wood assumes that action *from duty* is required and even possible only when duty and inclination diverge. This is a variant of the view that duty should serve as a 'back-up system'. His eventual solution of the present difficulty differs accordingly. See also the extensive discussion of this problem in his joint commentary with Schönecker, *Kant's 'Grundlegung'*, pp. 64–77.)

35 Moral philosophers too are much more likely to stick up for it; see IV 442 fn.

36 At the present level of everyday moral thought verging on philosophical ethics Kant roughly covers the same ground – that of strict and wide duty towards oneself and others – with his four well-known illustrations of the categorical imperative as later on in Section II.

37 The shopkeeper's case is discussed in the same paragraph as the classificatory scheme because Kant, perhaps rather too quickly, wants to get A-type action out of the way.

suggest not overcharging even very naïve customers. Yet the obscurity of motivation is essential to Kantian ethics,[38] and a direct inference from interest present to motive of action would be more than problematic. The mere presence of certain inclinations does not force a free agent to act accordingly; nor does the fact that, as in the case of a shopkeeper, one is justified in attributing an interest in certain actions warrant the pessimistic conclusion that the person in question acted, as Kant puts it, 'merely for purposes of self-interest'. Does Kant illicitly slide from the thesis that 'it is not nearly enough for us to believe that the merchant acted this way from duty and basic principles of honesty' (IV 397.26–7) to something more like: it is quite sufficient for us to believe that the merchant did *not* act this way from duty and basic principles of honesty? Alternatively, does he perhaps consider honest behaviour from motives of long-term commercial interest morally adequate as long as these actions reliably coincide with duty?

Both readings should be rejected. To understand the example of the shopkeeper we must bear in mind that Kant expressly calls him 'prudent', a word that jars in the present moralistic context ('prudent' is used in the same pejorative manner in the second *Critique*, V 35.31). Because a *prudent* businessman first and foremost cares for his own economic well-being, he almost by definition considers honesty the best policy, not an unrelenting moral requirement. Kant's pessimism with regard to the moral probity of shopkeepers does not rest on fallacious reasoning. It is simply an expression of his general scepticism about the actual moral quality of people's convictions, as opposed to the mere moral conformity of their actions, which is manifest throughout Sections I and II of the *Groundwork*.[39] Modern business life would confirm Kant's worst suspicions.

Why does an analysis of the acts of a prudent shopkeeper contribute little to our search for the moral law? It fails on two accounts. First, any individual case of A-type action, even if by accident it coincides with duty, is neither done *from* duty nor immediately morally valuable. It is good in so far as it serves some distinct purpose. Secondly, A-type scenarios like that of the shopkeeper, in which commercial interest reliably recommends the act that duty commands, are uninformative. Let us

38 The moral quality of an action is not an empirical property; see IV 406–8; and Kant is generally just as pessimistic about the fundamental moral attitudes (*Gesinnungen*) of human beings as he is optimistic about their moral cognitive capacities (see e.g. IV 407.1–16).

39 See also the thought in the moral philosophy lectures that people without much opportunity to indulge in vice such as an 'innocent girl from the country' are unlikely to be virtuous (*Collins*, XXVII 249.21–3).

grant Kant's assumptions that long-term self-interest invariably points him in the direction of treating his customers honestly, which is even more plausible in the age of paper trails and surveillance cameras than in the late eighteenth century.[40] If so, we have little reason to believe that we are observing an action done *from* duty when we observe a shopkeeper as he serves an inexperienced customer honestly. The previous paragraph announced that the nature of morality was to become apparent from inspecting the conflict between inclination and morality; and yet in the case of the prudent merchant self-interest and morality regularly concur. That is why the prudent shopkeeper will never present us with an unambiguous case of action *from* duty. His mettle will never be put to the test.

Furthermore, even if, Kant's pessimism notwithstanding, we encountered a shopkeeper of firm moral conviction, his actions would still in our eyes display the familiar regularities of prudential foresight, rather than characteristics of the less familiar moral law that in fact they spring from, and that we are trying to find. The regular coincidence of his prudent action with moral norms detracts from what is distinctive about the latter: its unconditional necessity. We are still lacking a plausible example of action from duty. That is why we now leave actions of type A behind. Rather ironically, Kant thought that the example of the prudent merchant, on which so much ink has been spilt, was unimportant.

¶ IV 397.33 By contrast (*dagegen*), there are cases in which people *ordinarily* act not for prudential reasons but from some more immediate interest. Moreover, Kant will soon stipulate the absence of natural inclination towards a 'dutiful' act, which helps us to identify actions that are clearly motivated by duty. It is only in such cases that the moral content of a maxim becomes apparent. When all other possible explanations of an action fail, we can be reasonably certain that it was done from an immediate interest in the moral action itself (C).

This so-called 'strategy of isolation'[41] is a heuristic device employed to reveal the distinctness of moral volition at a specific point in the

40 Yet in the lectures on moral philosophy there is the example of a dishonest shopkeeper and religious hypocrite (*Collins*, XXVII 332.14–17).

41 This is H. J. Paton's expression. He does, however, conceive of the task of the strategy somewhat differently. According to Paton, Kant wants to 'justify' the 'contention that a good will – under human conditions – is one which acts for the sake of duty'. For this purpose, we must 'judge whether they possess the supreme worth which we have described to a good will' (see Paton, *The Categorical Imperative*, p. 47). It seems to me that this 'contention' is assumed, rather than argued for, even if it is of course confirmed in the course of singling out actions that are clearly done from duty. The point of isolating actions from duty is rather to reveal the attitude that makes them so special.

argument of Section I. It was announced in the first paragraph of this subsection at IV 397.9–10 above. Kant does not wish to say that actions possess moral worth only when they are done without any concurrent inclinations, or even contrary to opposing inclinations; even less that these adverse conditions are desirable or that we should try to bring them about. He does, however, hold the view that there is something uniquely admirable about action that is not supported by, or perhaps even defies, strong inclination, but this should not be construed as a reason to provoke such conflicts.[42] It is only in difficult cases that the moral content of a robustly moral practical attitude becomes apparent. Within the consecutive train of thought of the three subsequent examples, an action is revealed to be morally good 'first' (*allererst*) when it is certain that it is no longer determined by unsuitable incentives (IV 399.26).

The *second* example, of the 'unfortunate man' who is tired of life, concerns a strict duty towards one's own person. For the sake of the argument, let us accept the view that there is a moral duty to preserve one's life. It is plausible to assume that, in addition, we normally have a strong direct inclination to do so. At this point, Kant's pessimism about actual moral attitudes is again tangible. He assumes, as he did in the first case, that human beings commonly subordinate duty to inclination; and he concludes that for the most part they preserve their life from immediate inclination, not from duty. It is only when someone has lost all interest in life – when, in other words, there is no other plausible explanation of the fact that he is preserving his life – that we may conclude that his behaviour is determined by firmly held moral principles. Again, Kant does not wish to say that *only* an unfortunate

42 This is the grain of truth contained in the 'battle citation model' R. Henson attributes to the Kant of the *Groundwork*; see his 'What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action', *Philosophical Review* 88 (1979), 39–54. Confirmation of Kant's high esteem for onerous moral action can easily be found elsewhere in his works. In the *Critique of Practical Reason*, virtue is said to 'reveal itself most splendidly in suffering' (V 156.31); in the *Anthropology*, he announces that there is not much in doing what is easy (VII 148.6–8); in a handwritten note, Kant says that the good will shines even more brightly 'on the black background of misfortune' (R 6968, XIX 216); and according to student notes of a lecture course delivered when he had just dispatched the manuscript of the *Groundwork* to his publisher, if there was a being 'whose good will not infrequently caused him damage' his good will 'would be all the more luminous' (*Mrongovius* II, XXIX 599.36–8); and, most explicitly: 'According to its moral worth, an action is worth the more the costlier it is for me' (XXIX 613). Kant evidently thought that there was something particularly admirable about moral action that does not just ignore the agent's inclinational state but directly thwarts his desires. If the reconstruction of the 'first proposition' given above is correct, an action possesses *some* moral worth if it is willed directly, and non-contingently, for the appropriate moral reason, i.e. if its maxim has moral content. Now it seems that the costliness of moral action can somehow *increase* its moral worth. But that is no reason to try to maximise it.

person who yet preserves his life out of duty acts in a morally worthy manner, or that we should try to make ourselves miserable to be able to try our maxims. More fortunate human beings can, and indeed should, similarly take care of their lives on moral grounds. Yet it is only when an unfortunate person still preserves his life when others would carelessly have thrown it away that the moral content of his maxim is revealed.

¶ IV 398.8 The notorious *third* case – the 'philanthropist' – concerns the duty of beneficence, a wide duty towards others. There are two distinct ways of helping others from inclination: first from some higher-order motive such as the expectation of future rewards or vanity, category A – perhaps one fancies oneself as a benevolent sort of chap and thus concludes that it would be a splendid thing to be beneficent when the occasion arises – and secondly from immediate sympathy, category B. As the former category has already been set aside,[43] we are concerned only with the latter kind of example: beneficence from immediate sympathy.[44]

Yet Kant again denies that any such action can possess moral worth. This is indicated by the fact that beneficent inclinations 'are on the same footing with other inclinations' (*[sie gehen] mit anderen Neigungen zu gleichen Paaren*, IV 398.15), i.e. they can subjectively be overridden by other inclinations.[45] They are therefore an unsuitable foundation for a necessary, unconditional moral command. The philanthropist's sympathetic desires can be 'overclouded' by sorrow (IV 398.21). Furthermore, the person in our example might have been naturally cold-hearted like those who feel no natural inclination whatsoever to help those in need (IV 398.27–34). The first of these two cases is introduced to demonstrate the vulnerability of immediate inclination in the life of an individual. If the 'fallen' philanthropist still acts beneficently, it must be from the motive of duty. The second illustrates the unevenness of the gifts of nature, which knows of no universal and necessary moral

43 Also, vainglorious actions are even less likely to be considered paradigms of moral goodness than those that spring from commercial interest.

44 Note that Kant declares being beneficent 'where one can' (IV 398.8) to be a duty. This seems to point to a fairly demanding overall picture of Kantian ethics. In a given situation, there is little room for inclination to decide whether to act on a moral maxim of practical love, despite the misleading double negative at IV 421 fn. There may, however, be moral checks and balances. See my 'Good but not Required? Assessing the Demands of Kantian Ethics', *Journal of Moral Philosophy* 2 (2005), 9–28. It is worth noting that the demandingness of morality is softened by the fact that the moral law is self-imposed.

45 For examples of the kind of trade-offs Kant has in mind see *Critique of Practical Reason*, V 23.22–9.

commands. It is also stricter, because in the first case the philanthropist was still in possession of his benign inclinations, despite the fact that his sorrows rendered them inactive. For the purpose of finding a clear example of action *from duty* it is safer to stipulate them away altogether, as in the example of the insensitive man who acts beneficently from duty.

If human behaviour were determined by inclination, beneficent action would be impossible whenever sympathy is either too weak or altogether unavailable. If acting from immediate inclination were an example of truly moral action, morality would be exposed to the whims of 'moral luck'. There could be no commands that are strictly universal and necessary. Only the philosophical stipulation of reverence as the motive of action from duty resolves the difficulty. It is at all times available to all agents because they carry the 'source'[46] of it within themselves, even to those who do not experience the slightest inclination towards the act externally judged to be morally right. That is how the moral quality of the 'fallen' philanthropist's maxim is revealed. If in the absence of, or even contrary to, strong inclination a person reliably still acts beneficently, it must be from a maxim of acting for the sake of duty, a maxim that possesses 'moral content'. It is no accident that the language of the last part of this paragraph is reminiscent of Kant's praise of the good will, as now expressed in actions from duty.

Kant's remarks suggest that an 'inclination to honour' (IV 398.15–16) could be a comparatively reliable substitute for the motive of morality. In a handwritten note, Kant offers us a hint as to why this is so. He writes that 'honour is the only inclination that can be based on principles, because the impartial applause of others rests on principles, which is why the love of honour is akin to virtue' (R 7215, XIX 287).[47] We want our actions to be approved by others. Universalisation is 'externalised' because *we* pay attention to the *disinterested* applause of others. Yet Kant does not think that morality can be *grounded* in the honourable. Caring for the impartial approval of others leads to action coinciding with the letter, not the spirit, of duty.

46 *Quell*, IV 398.35, see IV 405.25, IV 407.37 and IV 426.2. Kant is alluding to his theory of autonomy. The source of moral action is the human will as practical reason. See IV 431–3 and IV 440 below.

47 See also Kant's early essay on the *Feeling of the Beautiful and Sublime*, II 217, the *Metaphysics of Morals* VI 464 (*honestas externa*) and his lectures on moral philosophy (*Collins*, XXVII 408–9). As with happiness and the virtues, Kant hardly pays due attention to the highly moralised ancient (Stoic) conception of *honestas*. For a more sympathetic account see Kant's lectures on practical philosophy, *Mrongovius* II, XXIX 631.34–632.15.

¶ IV 399.3 Kant *fourthly* discusses not a wide, but rather at first a so-called 'indirect' duty towards one's own person: the duty to secure – not: to pursue! – one's happiness. An 'indirect' duty is not a duty *sui generis* or duty proper, which is morally binding in its own right. It is not a species of duty alongside the categories of perfect (strict) or imperfect (wide) duties, officially introduced at IV 421.21–3 below. It is, of course, a matter of duty to perform actions indirectly commanded by duty; but the injunction expressed by an 'indirect' duty commands only accidentally, as a mere (permissible, possible, appropriate) means to realising an end directly commanded. 'Indirect' duty is generated by technical rules, not by the moral imperative, and is thus essentially consequentialist in character. Complying with 'indirect' duty – performing a certain act required by instrumental rationality – does not possess moral worth as such. In the present example, the end is that of one's own moral integrity, which is in peril when one's happiness is 'under pressure from many anxieties and amidst unsatisfied needs' (IV 399.4–5).[48]

A reconstruction of the fourth example is further complicated by the fact that Kant appears to use 'happiness' in two subtly different senses of the word. There is, first, the purely formal 'canonical' Kantian definition of happiness as the continuous feeling that results from the satisfaction of the sum of one's inclinations, which requires prudential foresight. In the present passage, there is, secondly, a more normatively charged conception of happiness according to which, for example, being healthy is still part of happiness even if a person comes to lack all inclination towards his own health. We can consequently conceive of two different commands to look after one's own welfare in the face of temptation to neglect it. It is, first, indirectly a matter of duty not to become so poor that there might arise an overwhelming temptation to steal other people's belongings; and there is, secondly, the duty of the person suffering from gout of the foot (podagra) to take care of his own bodily well-being solely from direct duty. The problem is not that this person has lost the natural interest in his own long-term well-being. Rather, because of the uncertain success of physical, dietary or medical care, his happiness in the canonical formal sense no longer supports what

48 See *Metaphysics of Morals* VI 388.17–30 on the 'indirect' duty to preserve one's affluence, health and strength. The later work also contains other examples of similarly accidental 'duties' commanding actions only mediately morally relevant: compassion (VI 457.26), conscientiousness (VI 401.21) and behaviour resembling morality towards animals (VI 443.23). On 'indirect' duty in general see Appendix D.

is commonly perceived as the prudential course of action. It may be entirely rational for him – in the sense of instrumental 'means-ends' rationality – to prefer the pleasures of the moment to bodily health. In fact, he stands a good chance of maximising his overall happiness that way. Yet Kant now apparently accords caring for one's own health, as a conventional element of happiness, some direct moral weight,[49] in the way he later recommends developing one's natural talents (IV 423.13–16, cf. IV 430.10–17), quite independently of how this effects the sum total of one's inclinations or the moral dangers involved in neglecting one's well-being, which are not even mentioned in Kant's discussion of what the person with gout should do.[50]

The case of the person suffering from gout again confirms the 'principle of non-contingence', the most promising candidate for the missing first proposition. For an action to possess moral worth, its maxim must be such that the action that accords with duty occurs reliably and by force of necessity, without any inclination or even contrary to whatever the agent's inclination might dictate.

¶ IV 399.27 If inclination is morally irrelevant, the biblical commandment to love one's neighbour[51] must be interpreted as a command of practical charity from principles of duty.[52] Only practical charity can be sustained in the absence of benevolent sentiments, or even in the face of 'natural and unconquerable aversion' (IV 399.30–1) towards, for example, an enemy.[53] A mere natural sentiment, Kant thinks, can at best be cultivated over a certain period of time, but it cannot

49 That health is a necessary element of happiness is called into question at IV 418.19–21. The present argument might have been more persuasive if Kant had introduced other long-term interests to which the gout sufferer is committed but tends to neglect in his pleasure-seeking battle with the disease, such as wealth or knowledge.

50 Kant's theory of duties to the self might point to an intuitively much more adequate, if thoroughly moralised, account of human flourishing than his purely formal conception of happiness.

51 See Matthew 5, 43–4, referring back to Leviticus 19, 18; cf. Luke 6, 27; 6, 35. Note that Kant's realignment of scripture with his own moral philosophy is characteristic of his approach to morality and religion as a whole; see IV 408.35–7 for another good example. In the present paragraph, he does not say that 'this is what passages from scripture say', but rather that 'undoubtedly' it is in the way characterised above that 'we are to understand passages from scripture' (IV 399.27), *sc.* passages that are nonsensical otherwise.

52 On the difference between 'pathological' love – based on sensation, the passive part of human nature, i.e. inclination – and 'practical' love see also *Critique of Practical Reason*, V 82.18–84.21, where the biblical command to love one's neighbour (V 83.3–4) is discussed in very similar terms.

53 See Kant's lectures on moral philosophy, *Collins*, XXVII 413–14 and XXVII 417, and the second *Critique*, V 76.26, where the natural feeling of love is opposed to the rationally wrought incentive of reverence for the moral law.

immediately and unconditionally be commanded – which is precisely what moral imperatives do. Also, there are inclinations and aversions incapable of being overcome by any desire other than the motive simply to do what duty commands. Human beings are thus in charge of, and responsible for, the reflective principles on which they act, but not – or at any rate not directly – for their emotional reactions.

c. Proposition 2: The moral worth of an action does not lie in the effect intended but rather in its maxim [to be judged by the standard of a formal principle]

¶ IV 399.35 Kant now employs a strategy of exclusion to advance towards the idea of a purely formal moral principle. There are three steps. *First*, as the preceding examples show, one and the same apparent intention or purpose (*Absicht*) – such as that of preserving one's own life or of helping other people – can be expressive of different 'maxims' that vary with regard to their moral content. That is why, *secondly*, the moral worth of an action cannot consist either in an intended purpose or in its successful actualisation. (In today's philosophical terminology, Kant dismisses both subjective and objective consequentialism.) The moral quality of an action must rather be located in the maxim that first determines which purpose is to be intended, or whether an action is at all to be performed. Kant, *thirdly*, goes beyond the second proposition as explicitly stated. As all matter – all specific purposes that rest on subjective incentives – has been disqualified, positive moral value is thought to be characterised by the fact that a maxim is determined exclusively by the 'formal principle of volition as such' (IV 400.14), which is simply called 'the law' in the paragraph that follows. Note that we still do not know anything specific about this principle or law; or indeed whether human agents are capable of acting solely for its sake.

This is the general drift of Kant's argument for the second proposition. The details, however, are complicated by his ambiguous usage of the word 'principle' (*Prinzip*). It is not always clear whether, speaking of the 'principle of the will' or 'of volition', Kant is referring to the *subjective* principle of action on which the agent *does in fact* act – his maxim – or to the *objective* principle of morality – an ethical command, 'the law' – with which his actions *ought* to cohere.[54] The above reconstruction suggests

54 See Kant's footnote attached to the subsequent paragraph, where at long last the term 'maxim' is defined and distinguished from the practical law (IV 400 fn.).

he is initially referring to subjective principles and turns to an objective principle only towards the conclusion of the third step, at IV 400.14.[55] The phrase 'the formal principle of volition as such' clearly points to the objective moral law. It is only at that stage that the objective condition that a maxim must obey for an action to be genuinely moral is explicitly mentioned.[56]

Kant's discussion of the will at the Herculean 'crossroads' of motivation (IV 400.12) neatly illustrates the thesis that even a free will cannot be completely independent of determining laws. The decisive question is that of which *kind* of law one allows to determine one's will.

d. Proposition 3: Duty is the necessity of an action from reverence for the law

¶ IV 400.17 It is not easy to see how Kant intends the 'third proposition' to follow from the previous two – especially since he fails officially to call any proposition the 'first'. Furthermore, the cautious formulation of the opening sentence – 'I would express as follows' (IV 400.17–18) – suggests that Kant did not, as might otherwise seem natural, consider the inference as proceeding along the lines of a tidy syllogism with two premises and a conclusion. How, then, is the definition of duty as 'the necessity of an action from reverence for the law'[57] meant to be the 'consequence' of the two preceding propositions?

The first proposition – as presented above – left open the question of the conditions that must be fulfilled for an action to possess moral

55 Owing to the uncertainty surrounding the word 'principle' in this passage, H. J. Paton detects a distinction between 'formal' and 'material' maxims (*The Categorical Imperative*, pp. 61, 71 ff.). Yet I doubt whether this is what Kant had in mind. All maxims are material in the sense that they contain an end, even if their end is formal in the sense that it is primarily directed at the *willed* action itself as conforming to moral laws, not at a purpose or intention to be realised. The ambiguity of the word causes interpretative difficulties once again at IV 412 in Section II below, where the will is officially defined as the capacity to act in accordance with 'principles'.

56 The expression 'every material principle' at the very end of the present paragraph (*alles materielle Prinzip*, IV 400.15) refers to subjective incentives directed at the realisation of some specific object.

57 Kant uses the objective term 'necessity' (*Notwendigkeit*), not the subjective 'necessitation' (*Nötigung*). The official definition of duty does not therefore support Wood's thesis that acting from duty is only required when, owing to conflicting inclinations, it is difficult to do so (*Kant's Ethical Thought*, p. 43). Kant wants to say that when a command of duty applies, we have no choice but to act accordingly, and that reverence for a soon-to-be-disclosed law will serve as our motive for our moral action. Of course, objective necessity entails an element of necessitation in a finite will, but that is a matter of principle – because in morally relevant matters inclination is excluded from determining commands of the will – and not confined to single acts. Cf. the definitions of duty in the *Critique of Practical Reason*, V 80.25–9. See also Kant's 'Praise of Duty', V 86.22–33.

worth, i.e. for a morally correct action to come about reliably and independently of an agent's current state of inclination. Presenting the second proposition, Kant first located the moral worth of an action in its maxim, not in its purpose or surface intention, and then introduced the formality requirement that any morally valuable maxim must satisfy.

The basic idea of the present reconstruction is the following: the third proposition answers the first proposition's requirement of *non-contingent reliability* by referring to the criterion of *formality* that emerges in the discussion of the second; the thought of a purely formal law determining the will arouses in us *reverence* for this law – a novel notion, introduced only at this stage of the argument. This thought produces reverence, not a mere inclination, because it implies that *I* can freely determine *myself*, independently of any inclination, if I act to satisfy this law of reason, which is very much my own. (We catch a glimpse of Kant's doctrine of autonomy, officially introduced only at the end of Section II.) At the crossroads of motivation, reverence is the moral opponent of non-moral inclination. When the moral law speaks, reverence, like inclination, is available as an effective motive; but unlike inclination it is grounded in an objective, purely formal and universal law. If therefore the maxim of duty *subordinates* the satisfaction of inclination to reverence for the law, the morally right action non-contingently and necessarily follows. In short: within the formulation of the third proposition (i) 'necessity' points to the first, (ii) 'law' points to (a corollary of) the second proposition, and (iii) 'reverence' is the new element required to complete the definition of duty. *Voilà*![58]

Fn. IV 400 Kant finally distinguishes the maxim of an action as the (descriptive) principle that determines what an agent *does in fact* will from the law as the (normative) principle of what he *ought* to will (see also IV 420–1 fn.). If we were endowed with a holy will, which is always in perfect harmony with moral goodness, we would automatically – but still from rational insight and not just mechanically! – act in accordance with objective principles. Ascribing subjective principles (maxims) to us would be devoid of meaning. That is why in the *Critique of Practical Reason* Kant says that the concept of a maxim, as well as the concepts

58 The first and second propositions thus implicitly revolve around the two traditional attributes of the a priori: necessity and universality respectively. Alternatively, one might retain the conventional first proposition ('Action has moral worth only if done from duty') and leave the second proposition as it stands. They would still – in a weaker sense – point to the third proposition in the manner described above once their implications are clearly spelt out.

of an 'incentive' and an 'interest', can be applied to finite, imperfect beings only (V 79.27–35). By contrast, Kant's definition of duty rests on the motivational mechanisms of the human will. A person who does the moral thing from duty does so *from reverence for the law*, not because he or she wishes to perform the action in question anyway like a being endowed with a divine will.[59] Note that right from the beginning of the main argument (IV 397.1) Kant equates morally good volition with action *from* duty; and he considers the 'containment' of the concept of the former in that of the latter a matter of conceptual analysis. It can never be morally sufficient merely to obey the letter of the law but not to act in its spirit. We should, however, note that the import of acting from reverence for 'the law' is still unclear because the law in question has not yet been explicitly stated.

¶ IV 401.3 This new paragraph does little to advance the argument. Subjective consequentialism is explicitly rejected. If the moral worth of an action does not lie in its effect, Kant plausibly argues, it cannot consist in any principle that borrows its 'motivating ground' (*Bewegungsgrund*) from expected effects either, i.e. it cannot reside in a maxim that makes us act for the sake of some expected result. Moreover, effects can be brought about by powers other than the good will of a rational being, for example by some fortunate accident or purely by natural forces. They cannot therefore be *morally* good.[60] Kant again emphasises the peculiar character and dignity of a rational will and thus relates the results of the present discussion of action from duty back to the *Groundwork*'s opening line. Only action stemming from the representation of the law – which arouses reverence – is morally good, not the purpose or intended effect of an action – which may rest on a corresponding inclination.

Fn. IV 401 By contrast, the footnote attached to this paragraph is crucial for our understanding of Kant's theory of motivation. It is also easily misconstrued. The defence of 'reverence' as more than 'an obscure feeling' has done little to dispel the aura of mystery surrounding it. Its apologetic character is, however, witness to the fact that Kant was entirely aware of the philosophical extravagance of his moral psychology.

59 See the general definition of a will (below, IV 412.26–413.8), the idea of objective volition (IV 449.17) and R 7201, XIX 275.

60 This thought is strongly reminiscent of Kant's teleological arguments for the special character of the morally good will; see IV 394–6.

The fact that in the *Groundwork* we learn comparatively little about the exact psychological mechanism of action from duty is easily explained by its purpose.[61] Sections I and II are devoted to formulations of the principle that serves to discriminate good and bad moral actions, the *principium diiudicationis*; any discussion of moral psychology – the *principium executionis* – must be relegated to the footnotes. However, the results of these sections are hypothetical in that they depend on the existence of a moral incentive. That is why Kant returns to the question of why we take an interest in the moral law in Section III (IV 448–53). The problem of whether moral commands 'exist' or 'are real' cannot be discussed independently of the question of whether reliable moral motivation is available to the agent. Without our freely taking an interest in morality, categorical imperatives would not apply to us. As 'ought' implies 'can', they would not, in fact, be imperatives at all, but rather an idle wish, or a nice idea. The fact that our interest in morality turns out to be inexplicable is thus a severe limitation of the justificatory project of Section III.

The note naturally divides into three sections. In the first (IV 401.17–30), Kant emphasises the exceptional nature of the sentiment called 'reverence'. To be triggered, it does not depend on external objects of desire or aversion; and it is not, like incentives that express inclination or fear, directed at anything outside the agent's will.[62] Reverence is inevitably and reliably caused by the representation of the moral law in moral deliberation. The law – the universal law of morality – is then said to determine the will immediately. By contrast, inclination determines a specific law and urges the will to act on it, in order to obtain or realise some object. In short: reverence is a feeling that is effected by a law, whereas otherwise laws of action are effected by feelings, i.e. inclinations.

It is important to realise that at this stage Kant is referring to the will (*Wille*) in its *legislative* function. Reverence is not an epiphenomenon of morality, but rather that which *motivates* moral action. The motive that can give actions moral worth is neither fear nor inclination but 'simply reverence for the law', as Kant quite explicitly puts it at IV 440.5–7 below. These are actions *from* duty. A special incentive is still required

61 In the *Critique of Practical Reason* Kant devotes a whole chapter on the 'incentives of pure practical reason' (V 71–89).

62 'Inclination' (*Neigung*) is here used specifically for a positive leaning towards something, as opposed to the aversion or disinclination called 'fear' (*Furcht*). The latter is, of course, an inclination in the broader, more common sense of the word.

for translating the pronouncements of the legislative will into action, by means of the *executive* or 'elective' will (*Willkür*). Both functions of the will together form the will (*Wille*) in the comprehensive sense of the human desiderative faculty (*Begehrungsvermögen*). The representation of the moral law makes an incentive – reverence – available; and the agent must then consciously choose to act on this curiously immediate interest in moral action, not on inclination. It is only at that stage that his faculty of choice is determined in accordance with the command of the will. In short, Kant does not wish to say that in moral volition the will can do without any desire or incentive whatsoever because it is 'directly' determined by the law. An action from the motive of reverence is the most immediate way in which the executive will can elect to be determined by a law; but like any other action, an action from duty – from reverence for the law – still requires an incentive. The difference between both cases is that in action motivated by inclination the preceding judgement of what action would be best – of which law to act on – is determined by inclination, whereas in moral action it is independent of inclination by virtue of being freely determined by the moral law. We then – inexplicably, whether we like it or not – take an interest in moral action. As reverence arises reliably and is always at least in principle strong enough to motivate the agent to act morally, the question of how the moral law can motivate is, in the *Critique of Practical Reason*, equated with the question whether the human will is free (V 72.21–4, cf. IV 458.36–459.2 below).

Secondly, Kant describes the effect that reverence has on the self (IV 401.31–5). This brief psychological account is an excellent example of the way in which he sometimes slides back and forth between two different notions of the self: a 'lower' self of the agent, which seeks to realise his inclinations, and a 'true' or 'higher' self, reason. Reverence is grounded in the representation of the dignity of humanity in so far as it is moral, which infringes upon an agent's self-love because it disregards all specific ends that his inclinations urge him to realise. Yet reverence is generated by the fact that we recognise that, emphatically, *we* impose this law onto *ourselves*, of our own free will, and yet by necessity.

Thirdly, all so-called moral interest is declared to be reverence for the law (IV 401.35–40). The object of reverence is the law within the agent, and equally in all other moral agents. (Non-rational beings or things do not, like persons, inspire reverence. We cannot have any duties to them; see V 76.24–7.) That is why reverence is the incentive of both moral

action towards one's own person as well as morally good actions to others.

4 The law of duty, general conformity to law as such, is the condition of a will that is good in itself

¶ IV 402.1 The philosophical analysis of the concept of duty, which was meant to elucidate a generally held view about the value of good willing under human conditions (IV 397.6–10), has led to an important intermediate result. We now know that when a person acts from duty, it is objectively the moral law and subjectively reverence for the law that determines his will. This is precisely what makes volition and action unconditionally good. However, Kant's analysis is still incomplete because we do not yet know what kind of law this is, or what exactly the moral law commands. Philosophy now attempts, for the very first time, to answer these questions. Kant presents us with a first version of 'the highest principle of morality', which in the more technical second section of the *Groundwork* is officially christened the 'categorical imperative' (see IV 421.6–8).

The basic thought that underlies Kant's argument is the following. In order to realise any specific end we must follow commands that make use of certain laws (later called 'hypothetical imperatives'; see IV 414.23). If, for instance, I intend to make a cup of coffee I must grind beans, put the ground coffee in my cafetière, add hot water, stir the mixture, and so forth. I must pay attention to the specific empirical laws that enable me to realise my end: a cup of coffee. However, the preceding analysis of the concept of duty has shown that morally good action is precisely *not* action that is performed for the sake of some particular end that one intends to promote, realise or bring about. If so, all laws that apply only on condition that one intends to realise some such end must be excluded from our list of candidates. They cannot be the law that inspires reverence and motivates morally good action.

However, we have also seen that the will must be governed by something – by some law – to be determined to action at all (IV 400.13). The reason for this is that the will is a type of causality, a way of producing effects, and this must happen in a law-like fashion. Duty and morality would otherwise be no more than a phantasm. How could there be an action not determined by some law or other? As all *specific* laws have been discarded, the only candidate that remains for the position

of a moral law is 'the universal conformity to law as such' (*die allgemeine Gesetzmäßigkeit der Handlungen überhaupt*, IV 402.6, cf. IV 421.2–3). Kant spells this out as follows: one morally ought to act in such a manner that one can will the principle on which one acts to become a *universal* law. The example that follows illustrates how this novel criterion is to be applied.[63]

¶ IV 402.16 Consider a consciously false promise. Kant and common moral opinion agree that it is immoral, even in difficult circumstances, to promise to do something if one has no intention of keeping the promise. Unlike matters of prudence, moral questions permit of a clear and unambiguous solution.[64] There is no need for reason to engage in complicated and ultimately uncertain empirical research because it contains the a priori criterion of moral, non-consequentialist volition within itself: is it possible for me to will that the subjective principle under scrutiny – to escape difficult situations by consciously making false promises – is observed not just by myself, but also universally adopted by everyone else? I cannot. For if all other agents similarly subscribed to this principle, the very foundation of all promises would be undermined. If so, I would not be in a position to realise the end of the promise – to escape my present difficulty – by making a *false* promise. To put it somewhat differently, I must rely on other people's general good faith in truthfulness in situations like the present, and a fortiori on their faith in mine. I must certainly rely on the faith of the

63 All laws as such are universal, but only one is universal '*überhaupt*' in that its scope is not restricted to specific agents. Excluding all such specific laws, Kant is left with conformity of (the maxims, i.e. subjective principles of) actions with universal law as such, which quite legitimately he equates with a very first formulation of the categorical imperative. For example, the described regularity governing the preparation of coffee has normative force for anyone who wants to make a cup of coffee with the above utensils. It is rational for me to act on this principle if and only if it can equally be willed by any such person. The subjective rules that correspond to specific instrumental laws are universalisable in this sense. What would happen if everyone who desired a cup of coffee acted in accordance with the same rule? Everyone would get a cup of coffee! The moral law must be universal in a different sense. If, as a consequence of the 'third proposition', we discard all conditional laws from our search for the standard of morality, we arrive at the idea that the principle that correctly describes our action is legitimate only if it can be willed universally by everyone, regardless of his or her particular purpose. The rule in accordance with which I act must not conform to some particular law instructing me how to realise a certain purpose, but to law-like regularity as such. The question is then whether I can will the principle to which my proposed action commits me to be willed by everyone. (Of course, whether this thought experiment yields any substantive results remains to be seen.) The alleged 'gap' in the derivation of the categorical imperative will be discussed in more detail in a note on the parallel passage at IV 420.24 below.

64 This is a constant theme in Kant's practical philosophy; see e.g. the second *Critique*, V 36–7.

person I am trying to persuade – and it is morally wrong to apply double standards.[65]

¶ IV 403.18 Kant argues that in moral – as opposed to prudential – matters common human reason possesses, in the shape of the above formal law, a clear criterion that is easily applied. The question why human beings cannot but accord this law a special estimation far beyond the value of the objects favoured by inclination is an advanced philosophical question and therefore falls outside the scope of Section I.[66] The necessity of an action from reverence for this law is duty, the condition of a human will that is essentially good beyond everything else. By referring back to the first sentence, Kant indicates that we have reached the conclusion of the first 'transition' within the *Groundwork of the Metaphysics of Morals*. The concept of the good will is connected with that of the law through the concept of duty.

5 *Concluding remarks: common and philosophic moral cognition of reason*

¶ IV 403.34 The concluding remarks of Section I complete the transition from common moral thought to moral philosophy. We have been presented with an explicit statement of the rational *principle* of ethical thought, ever present as it may be in everyday moral judgement; and as investigating the a priori principles of things is a genuinely philosophical task,[67] we now firmly stand on philosophical territory. A much more technical investigation of the supreme principle of morality will soon follow in Section II.

Kant's enthusiasm about the new principle is striking. Considering the controversies that still surround Kantian ethics, we may have our doubts whether with the 'compass' of the categorical imperative in hand it is really so very easy always to decide whether an action is right or wrong. We should, however, recall that Kant is not trying to develop an ethical theory that, like Sidgwickian utilitarianism, is the prerogative of a philosophical elite. Rather, he is explaining a practice in which all rational agents make their own decisions, according to a principle they

65 The example of the lying promise is taken up again, and more exhaustively discussed, in Section II (IV 422.15–36 and IV 429.29–430.9).

66 It is taken up again in Section III, where we are still told that we lack *insight* into the source of reverence (IV 459.32–460.7). But Kant at least provides an account of the best possible explanation of the origin of the interest we are bound to take in morality.

67 See explicitly the third *Critique*, V 174.3–5.

naturally find within themselves. In his essay on *Theory and Practice*, he similarly asserts that in moral matters everyone is a 'professional' (*Geschäftsmann*, VIII 288.33).[68]

Kant notes a significant difference between theoretical and practical reason, already mentioned briefly in the Preface, IV 391.20–4. When *theoretical* reason leaves experience behind to venture forth into the field of the metaphysical it gets entangled in illusions and contradictions that can only be resolved by means of a 'critique' of pure (theoretical) reason (see A 2/B 6 ff., A 293/B 349). By contrast, *practical* reason flourishes when it abstracts from the world of sense and experience.[69] It is true, under the influence of inclination the powers of practical judgement can go astray and try cleverly to obstruct rightful claims – for example, the claims of conscience – as a pettifogger does in a court of law (*schikanieren*, IV 404.21).[70] Yet common practical understanding also takes great delight in subtle critical investigations of the moral worth of agents and their actions. That is why Kant recommends discussing the moral merits of individual agents and their actions as an educational exercise.[71] The academic philosopher's assistance is not needed. More likely than not he, or she, would only complicate things.[72]

As Kant notes towards the end of the paragraph, his optimistic assessment of the indigenous powers of human practical reason raises the question why moral philosophy is necessary at all. On the most natural reading, the following two paragraphs present a retrospective defence of the transition made in Section I, which takes common moral consciousness as its starting point and puts it 'on a new path of investigation and instruction' (IV 404.35–6). However, the following thoughts contain perhaps a hint of why philosophy cannot rest content with what has been established so far. We need to go beyond the exposition of an ethical system and its principles to settle the 'natural dialectic' of

68 Kant's appeal to Socratic practice at IV 404.4 is problematic. He is probably referring to Plato's *Meno*, where Socrates teaches a slave boy entirely ignorant of geometry to double the area of a square, which is meant to demonstrate the a priori character of geometry and the pre-existence of the human soul; or, alternatively, to Socrates' professed method of midwifery in the *Theaetetus*. For the most part, however, the more Socratic of Plato's dialogues follow the pattern of Socrates' cleverly exposing the ignorance of his interlocutors. On the Socratic method see also *Metaphysics of Morals*, VI 411.34 and VI 479.10. In the latter passage, the method is declared unfit for the education of young children, who first need to learn their moral catechism.

69 This capacity to distinguish pure and empirical reasons for action is elegantly illustrated in the 'calcareous earth' example: *Critique of Practical Reason*, V 92–3; see also V 36.3–8.

70 For the 'sophistry' of self-love in the moral court of justice see *Collins* XXVII 359.8–28.

71 On the proper and improper use of examples in ethics see IV 408–9 below.

72 See *Critique of Practical Reason* V 155.12–22, where Kant says that only philosophers can call the validity of moral norms into question.

practical reason mentioned in the subsequent paragraph (IV 405.13), and to reveal the 'source' of the principle the analysis of Section I has revealed (IV 405.25). This points in the direction of Kant's theory of autonomy, developed in the course of Section II.

¶ IV 404.37 A philosophical ethics is needed after all because practical reason has its own dialectic, i.e. a fatal tendency to get caught up in contradictions. Such contradictions occur once we take *empirical* factors into account – as opposed to the dialectic of *pure* theoretical reason. Inclinations, which are rooted in the physical side of human nature, raise illicit claims and try to seduce the agent into quibbling his way out of the strict commands of reason (for example, along the lines of 'This action is unlikely to cause any harm!' or 'One more won't hurt!'). A pure moral philosophy can remedy the situation by grounding moral laws more firmly in the human will. In the essay on *Theory and Practice*, Kant similarly suggests that it is the proper task of ethical theory to emphasise the purity and strictness of moral commands; and Section II ends on the metaphysical note of the inspiring and elevating *dignity* of a self-sufficient human will (IV 434.20 ff.), which is here said to be threatened by the claims of inclination. Discussing the various permutations of the categorical imperative at the end of that section Kant also returns to the topic of the practical use of philosophical ethics (IV 437.1–4).

¶ IV 405.20 For Kant, a 'critique' (assessment of the powers) of reason consists of two main tasks: the justification of the relevant synthetic judgements a priori, as well as the resolution of contradictions inherent in reason, in a certain mode of employment. As far as practical reason is concerned, in the last paragraph of Section I, and similarly in the Preface of the *Critique of Practical Reason*, Kant focuses on the latter question, whereas Section III of the *Groundwork* represents his attempt to answer the former. It is significant that with regard to (pure) *theoretical* reason the two tasks overlap: Kant proceeds to dispel the illusions and contradictions of speculation by distinguishing the synthetic judgements a priori that can be substantiated from those that cannot. However, according to the *Groundwork* pure *practical* reason is not 'dialectical', and the two tasks are separate in the practical sphere. That is why on the one hand Kant says that *pure* practical reason does not stand in need of a critique, and yet on the other applies himself to the question of how the categorical imperative as a synthetic principle a priori is possible within the framework of a partial, rudimentary 'critique of

pure practical reason' in Section III. Again, the revelation of the *source* of the moral principle is accorded to a metaphysics of morals (see IV 392.20 above). It turns out to be the will itself. The supreme principle of morality is a principle of *autonomy*.

BIBLIOGRAPHY

Ameriks, Karl. 'Kant on the Good Will', in *Interpreting Kant's Critiques*, 193–211 (Clarendon Press, 2003; first published in 1989)

Baron, Marcia. *Kantian Ethics almost without Apology* (Cornell University Press, 1995)

'Acting from Duty', in *Immanuel Kant. Groundwork for the Metaphysics of Morals*, trans. Allen Wood (Yale University Press, 2002), 92–110

Benson, Paul. 'Moral Worth', *Philosophical Studies* 51 (1987), 365–82

Henson, Richard. 'What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action', *Philosophical Review* 88 (1979), 39–54

Herman, Barbara. 'On the Value of Acting from the Motive of Duty', in *The Practice of Moral Judgment* (Harvard University Press, 1993), 1–22 (first published in 1981)

Johnson, Robert. 'Expressing a Good Will: Kant on the Motive of Duty', *Southern Journal of Philosophy* 34 (1996), 147–68

Korsgaard, Christine. 'Kant's Analysis of Obligation: The Argument of *Groundwork* I', in *Creating the Kingdom of Ends* (Cambridge University Press, 1996), 43–76 (first published in 1989)

'Two Distinctions in Goodness', in *Creating the Kingdom of Ends* (Cambridge University Press, 1996), 249–74 (first published in 1983)

Potter, Nelson. 'The Argument of Kant's *Groundwork*, Chapter I', in *Kant's Groundwork of the Metaphysics of Morals*, ed. P. Guyer (Rowman & Littlefield, 1998), 29–49 (first published in 1974)

Rickless, Samuel C. 'From the Good Will to the Formula of Universal Law', *Philosophy and Phenomenological Research* 68 (2004), 554–77

Schönecker, Dieter. 'What is the "First Proposition" Regarding Duty in Kant's *Grundlegung*?', in *Kant und die Berliner Aufklärung*, ed. V. Gerhardt, R.-P. Horstmann and R. Schumacher (De Gruyter, 2001), vol. III, 89–95

Stocker, Michael. 'The Schizophrenia of Modern Ethical Theories', *Journal of Philosophy* 73 (1976), 453–66

Stratton-Lake, Philip. *Kant, Duty and Moral Worth* (Routledge, 2000)

Timmermann, Jens. 'Good but not Required? Assessing the Demands of Kantian Ethics', *Journal of Moral Philosophy* 2 (2005), 9–28

Walker, Ralph C. S. 'Achtung in the *Grundlegung*', in *Grundlegung zur Metaphysik der Sitten*, ed. O. Höffe (Klostermann, 1993), 97–116

Section II: Transition from popular moral philosophy to the metaphysics of morals

To reform moral philosophy, Kant makes a fresh start at a more technical level. He proceeds, again by means of conceptual analysis, via one familiar and three (or so) new formulations of the principle of morality, from a general characterisation of the will to a discussion of the concept of autonomy. There are implicit references to 'popular' moral philosophy throughout. It is criticised for failing to distinguish between the pure and the empirical elements of volition, and for its confused attempts as a result to derive ethical theory from examples, i.e. ethical role models.

The main argument of the section starts with Kant's definition of the will. Via the distinction of various types of volitional faculty, and the various types of imperatives that address a human will, this leads to the general formulation of the categorical imperative, familiar from Section I. In the Preface, Kant declares a general definition of the will an unsuitable foundation of moral philosophy. Distilling the morally relevant parts of human volition he now demonstrates that he can succeed at a task his predecessors did not even realise was necessary. A first variant – the 'law-of-nature' formula – follows. It takes up the Stoic ideal of 'living in accordance with nature', which was still fashionable amongst Kant's late eighteenth-century colleagues. Similarly, the second variant uses the definition of the final good as expounded by teleological ethical systems. The third alludes to the Leibnizian idea of a 'kingdom of grace'.

These variants are not separate, independent principles – there can be only one supreme principle of morality. They rest on analogies that serve to bring morality closer to intuition. Also, at the level of philosophical rhetoric, they are *argumenta ad hominem*, gracious compromises that Kant offers to his opponents and their followers in the hope of winning them over. He intimates that these variations preserve a grain of structural truth contained in heteronomous moral systems, even if interpreted along traditional lines they are thoroughly mistaken. They are based on an idea of pure practical reason, but owing to the confusion of pure and empirical elements this idea is not sufficiently clearly developed. In the *Critique of Pure Reason*, Kant famously says that it is 'not at all unusual' to find that we understand a philosophical author 'even better than he understood himself' (A 314/B 370). The variants of the categorical imperative are cases in question. However, the three

formulations also mark a philosophical progression towards the central Kantian notion of autonomy, which would have to take pride of place in a later metaphysics of morals. The special value of moral beings, which has the power to inspire us to moral action, first comes into focus in the derivation of the second variant. Ultimately, the human will is revealed to be the sovereign source of the supreme principle of morality.

1 Preliminaries

a. The origin of the concept of duty is not empirical but a priori

¶ IV 406.5 The concept of duty analysed in Section I is grounded in common moral thought but not therefore, Kant warns us, in anything empirical. Ordinary morality and experience should not be confused. Experience is silent in matters of moral value; it teaches us only how things are, not how they ought to be. Also, it is impossible empirically to decide questions of moral worth, i.e. whether an action was done from inclination or solely from a sense of duty. Neither the agent nor an external observer has any conclusive evidence whether the maxim of an action possesses or lacks moral content. All observable actions follow the natural laws of cause and effect. Even the closest empirical scrutiny of human behaviour will reveal inclination-based motivation only. Moral worth does not surface.

It follows that it is less than obvious that there is, or could be, a moral action at all. Empirical investigation rather confirms the suspicion that all action is an expression of the agent's inclinations, and thus in some rarefied sense 'egoistic'. In that case the analysis of the concept of duty in Section I, qua analysis, would still be correct, although the thought that human beings might ever act from duty would be fantastical. As ought implies can, there could be no instance of a moral command if the human will could be shown to be determined by the forces of nature. Kant takes this danger very seriously. It is due to the elusiveness of morality that philosophical sceptics, past and present, challenge the reality rather than the meaning of the concept of duty.[1] The binding nature of moral norms is the central problem of Section III. How can moral concepts have any significance for us?

1 See IV 441–3 below. Epicurus is frequently criticised for this explicit eudaemonism; see *Critique of Practical Reason*, V 24.15–25 and V 40–1. However, at this point Kant probably has Garve in mind, whose commentary on Cicero's *De officiis* is said to have inspired him to write a foundational treatise on moral philosophy (see Introduction above).

¶ IV 407.1 What counts – in morality, as opposed to the legal matters – is not visible actions but the inner principles of actions that are invisible, i.e. the agent's maxims. The shopkeeper's behaviour does not reveal whether he is thoroughly honest or simply a prudent businessman. But even when it is less obvious that prudential interest or some more direct inclination support the right act, neither a spectator nor the agent himself can ever be entirely certain to have witnessed a truly moral action.

We learn more about the reasons why maxims are not accessible empirically or introspectively in the final section of the *Groundwork*. Experience is limited to natural events, but the maxims on which human beings act are adopted freely, i.e. *independently* of natural causation. In fact, they are the grounds of the observable character of human action. Moreover, the dialectic inherent in human practical reason makes us vulnerable to self-deception. We flatter ourselves that our principles are more moral than in fact they are. An action that coincides with the letter of the law may well be the result of a maxim based on inclination.

The sensual side of our nature urges us to maximise the realisation of our inclinations: we would like to be happy. We naturally put inclination first when duty is silent; and it is very tempting to prefer inclination even when it runs counter to morality. That is why inclination opposes a maxim that makes its satisfaction conditional on moral permissibility, despite its consequentialist tendency to be concerned about outcomes, rather than acts of volition. In cases of conflict, inclination urges us to reason our way out of the strict commands of duty for the sake of occasional 'exceptions' from laws that the agent at the same time must will to stay in place (see IV 424.15–20) – again an expression of the tendency to consider oneself more moral than one actually is.

¶ IV 407.17 Making duty an empirical concept advances the cause of the moral sceptic. Kant parenthetically alludes to the categories of theoretical reason, or 'pure concepts of the understanding' (some of the 'other concepts', IV 407.21), whose characteristic universality and necessity cannot be explained empirically either. As with the concept of duty, empiricist philosophers are likely to give up on the claims that traditionally accompany our use of the categories, and proffer an error theory instead.[2] In short, both in the theoretical and in the practical

2 See Kant's reply to Hume, *Prolegomena*, IV 257–8.

sphere it is tempting to abandon claims that can be substantiated only by an argument revealing the way in which they are grounded in synthetic judgements a priori.

Kant is characteristically optimistic that most of our actions, though not genuinely moral, at least coincide with duty. This is due in part to the regulations of a well-ordered society and to human nature.[3] However, an experienced moral observer quickly detects the wiles[4] of self-love that make moral action difficult. Our 'dear self'[5] tries to dislodge the just claims of reason, and substitute the claims of inclination instead. *Prima facie*, all this is bad news for the moral philosophical cause. Yet Kant makes a serious appeal not to give up faith in the possibility of genuinely free and moral action. He draws a uniquely sharp distinction between facts and norms, or 'is' and 'ought'.

¶ IV 408.12 As experience is capable of revealing contingent truths only, our conviction that all moral laws are universally and necessarily valid for all rational beings cannot be justified empirically. They must, if at all, be grounded in pure practical reason.[6]

b. On the limited value of exemplars in ethics

¶ IV 408.27 What kind of examples does Kant have in mind when he dismisses *Beispiele* as overrated in popular moral philosophy? Ostensive examples, held up as ideals; in other words historical, literary or biblical role models and the allegedly exemplary conduct they display. He is not concerned with hypothetical examples, such as thought experiments or illustrations (although by virtue of being illustrations of something they can hardly be foundational in ethical theory either).[7] In accordance with a distinction proposed in the *Metaphysics of Morals*, Kant could have used the word *Exempel* (exemplar), rather than *Beispiel* (instance) to avoid confusing his readers. The former expression refers to the way

3 In the present passage, Kant concedes this conformity to duty 'from love of humanity' or 'philanthropy' (*Menschenliebe*, IV 407.23), which notoriously fails to qualify as a moral sentiment (see e.g. the essay on the 'Right to Lie', VIII 425–30).

4 *Tichten* [sic] *und Trachten*, or 'scheming and striving' (Paton), a common expression at the time that echoes Luther's translation of Genesis 6, 5 and Isaiah 59, 13.

5 Note that the 'dear self' associated with inclination in this passage is not our 'true' self, which is reason; and that, according to Kant, only the former, by virtue of its lack of universality, is selfish.

6 Kant echoes what was said about the necessity of pure moral philosophy in the Preface; see IV 388–91.

7 For this distinction see O. O'Neill, 'The Power of Example' in *Constructions of Reason* (Cambridge University Press, 1989). On the general topic of examples in moral education see R. B. Louden's 'Go-carts of Judgment. Exemplars in Kantian Moral Education', *Archiv für Geschichte der Philosophie* 74 (1992), 303–22.

in which we 'take something [or someone] as an example', which is the meaning required in the present context. The latter is that which is brought forward to 'clarify an expression' (VI 479 fn.).[8]

Kant rejects an ethics based on exemplars, favoured by eighteenth-century popular philosophy (see IV 412.17–18) as well as neo-Aristotelian proponents of 'virtue ethics' two centuries later, because any example implicitly points to a standard presupposed in our judging it exemplary in the first place. Kant objects to the fact that these implicit standards are never acknowledged to be primary, clearly stated or explicitly defended.[9] Similarly, in the lectures on moral philosophy he finds fault with parents who point to the behaviour of the neighbour's child as exemplary. Kant thinks that this is likely to breed envy and resentment, rather than virtue. A child thus admonished will think: 'If only there was no Fritz across the road to be compared to, *I* would be the best child in the neighbourhood!'[10] We once again lack a proper, non-relative standard.

The Bible encourages us to follow Christ's example.[11] Yet even he, the 'Holy One of the Gospel', needs to be judged by an explicit principle for us to recognise him for the perfect role model that indeed he is.[12] As usual, Kant reinterprets a passage from scripture (Mark 10, 18; Luke 18, 19; cf. Matthew 19, 17) in the light of his own moral theory (see,

8 There is nothing wrong with using fictitious illustrations to clarify a principle; and as they are precisely that, there is little temptation to regard them as foundational in moral theory or practice. Note also that in Section I Kant put non-illustrative examples of concrete exemplary action to proper use by following up implicit hints that pointed in the direction of the principle of morality.

9 Note that Kant's objection does yet entail that an exemplary person consciously applies a certain practical standard; that requires an additional argument.

10 *Collins*, XXVII 437.18–438.20; see also XXVII 359.29–39. In the first *Critique*, examples are dubbed the 'go-cart' or 'walker' (*Gängelwagen*) of judgement (A132–4/B171–4). They never completely correspond to the principles of e.g. medicine or jurisprudence and are therefore an imperfect substitute for those who lack the power of judgement. There can be little doubt that in the early twenty-first century our 'go-cart of judgement' is the bullet point.

11 See e.g. Paul's first letter to the Corinthians 11, 1.

12 Kant often praises the New Testament for its purity and perfection, e.g. in the lectures on moral philosophy, *Collins*, XXVII 301.28–9. The role of Christ as a perfect exemplar is discussed at length in *Religion*, VI 54–60; see *Collins*, XXVII 250.14–20. Kant's target regarding the use of religious examples in the present passage may be A. G. Baumgarten. See §§ 133–9 of his *Ethica Philosophica* of 1763 (XXVII 904–6) and the corresponding sections in *Collins*, XXVII 332–4, which is very close to the *Groundwork* in spirit and wording. Kant there explicitly distinguishes 'following after' an example (*Nachfolge*) and 'imitation' (*Nachahmung*); see also XXVII 322–3. Kant occasionally recommends the former but he rejects the latter throughout; see e.g. *Anthropology*, VII 293.3–13. The difference is that in *following* an example one acts for the same good reasons as the person in question, whereas *imitating* a person is founded on the mere wish to be like that person. The necessity of the moral law as the original (*Urbild*), guideline (*Richtmaß*) and model (*Muster*) is also emphasised at XXVII 294.31–3. On the relation between ideal and example see *Critique of Pure Reason*, A 569/B 597 f.

for example, IV 399.27–34). In the *Conflict of the Faculties*, he defends his conviction that biblical exegesis must always be guided by ethical principles. Morally repugnant doctrines like Paul's theory of predestination are to be explained away (see VII 41.14–22, VII 66.18–67.2). Moreover, as we learn from the *Metaphysics of Morals* and the second *Critique*, moral education must precede religious instruction (VI 478.29–35); and all trustworthy evidence for God's existence is essentially moral (V 124–32). Calling God 'the highest good' at IV 409.1 Kant might – somewhat casually – be referring to the perfection of God's good will, in the spirit of the *Groundwork*'s opening lines, but he also deserves this epithet because the realisation of the highest good in the conventional sense, the union of morality and happiness, requires divine intervention.[13]

In the present passage, Kant argues for the ethical priority of the standard implicit in moral examples: the moral law. It is, however, important to note that his attitude towards historical or fictional examples in moral education – rather than ethical theory – is not wholly negative. We have already heard about the interest pure practical reason takes in assessing particular cases of morally good or bad action (see IV 404.16–28). The famous hardened villain of IV 454.21–2 is a case in question. When presented with examples (!) of decent conduct, he cannot but approve and wish that he were thus disposed, however painful it may be for him to become a better man. This theme is further developed in the Methodology of the *Critique of Practical Reason*. Discussing the moral worth of 'this or that action' is advocated as a kind of useful party game which shows human reason at its most subtle and best. Kant recommends that the case of a decent man who fends off temptations and threats and refuses to betray an innocent Anne Boleyn be held up as exemplary in the moral education of children. In accordance with the gradually changing nature of Henry's incentives, Kant describes four successive stages of observatorial reaction: approval, admiration, amazement, finally adoration and the fervent wish to be able to be such a person oneself (but naturally *not*, he fortunately hastens to add, to be in that person's predicament, V 156.18–19). It would seem that in the course of the 1780s, the importance of examples is strengthened by Kant's increasing faith in the motivational powers of a morality given to us as a 'fact of reason'. Of course, the purity and

13 See the distinction between God as the highest 'original' good and the best world as the highest 'derivative' good in the second *Critique*, V 125.22–5. The two readings are compatible; see V 131 fn.

authority of human autonomy and its standard, the moral law, must be emphasised throughout.

c. True and false popularity in moral philosophy

¶ IV 409.9 This paragraph is complicated by a host of counterfactual negations, some superfluous by the standards of modern grammar. Kant is saying that if there were no principle of morality resting on pure reason, there would be no need to proceed with his philosophical project of grounding a metaphysics of morals – but of course there is such a principle. Moreover, as there are currently so many adherents of a popular moral philosophy who confuse the pure and the empirical, we must give it a try (see also *Critique of Practical Reason*, V 24.25–31).[14]

¶ IV 409.20 Rather polemically, and amusingly, Kant again apologises for the inaccessibility of some parts of the *Groundwork* (see Preface, IV 391–2). Pure moral theory must come first, careful popularisation second. One is reminded of Cicero's *De officiis* and Christian Garve's *Philosophische Anmerkungen und Abhandlungen*, both equally eclectic and riddled with historical examples. It is difficult to believe that at least the present attack on contemporary popular moral philosophy was not inspired by Garve's 1783 twin publications. However, like G. A. Tittel, the *Groundwork*'s first commentator, Kant's modern readers may well be wondering whether his diatribe was really dictated to him by *pure reason*.[15]

¶ IV 410.3 Kant now accuses the different schools of thought in popular moral philosophy of arbitrarily confusing ethical principles, mostly supposed to be derived from facts about human nature. The project of a metaphysics of morals is alien to these populists. In particular, Kant is alluding to David Hume ('the special vocation of human nature'), the Stoics ('the idea of rational nature as such'), Christian Wolff and his school ('perfection'), Christian Garve and other eudaemonists

14 The exact wording of the last clause (IV 409.18–19) is uncertain. Instead of the second edition's 'preponderance' (*Übergewicht*), the first edition has 'truth' (*Wahrheit*), which makes no sense whatsoever. It is likely to have been 'choice' (*Wahl*) in Kant's original manuscript.

15 G. A. Tittel, *Über Herrn Kant's Moralreform* (Pfähler, 1786), p. 25. Tittel was following in the footsteps of his teacher, J. G. H. Feder, who had edited Christian Garve's review of the *Critique of Pure Reason* for the Göttingen review. In a letter dated 11 June 1786 Kant's friend J. E. Biester calls him the 'weak shadow of the weak Feder' (X 457.25–6, No. 275 [255]).

('happiness'), Francis Hutcheson ('moral feeling') and Christian August Crusius ('fear of God').[16]

Fn. IV 410 Kant emphasises the primacy of pure as opposed to applied moral philosophy. The latter involves empirical – in fact psychological – knowledge but is still rooted in metaphysical laws, as the term 'applied' indicates. On the notion of 'applied' as opposed to 'empirical' logic, see IV 387.17–21 above.

d. The primacy of metaphysics in moral philosophy

¶ IV 410.19 Kant returns to the theme of the Preface (IV 389.36). A metaphysics of morals does not just satisfy our philosophical curiosity, it also helps to root moral principles in people's hearts by clearly presenting to them the dignity of a will that is free to comply with moral laws even in the face of the strongest opposition by inclination. The second section, which ends on the note of autonomy, prepares the way for this – it is something a non-pure, mixed moral theory cannot achieve.

The present paragraph contains a hint of Kant's doctrine of moral awareness, which is more fully developed in the second *Critique*. The consciousness of one's moral obligation is the *ratio cognoscendi* of human freedom (V 4 fn.). Similarly, the pure representation of the moral law is said to effect that reason 'first becomes aware that it can of itself also be practical' at IV 410.28–9, i.e. it first becomes aware of its autonomy as *pure* practical reason. Characterising the views he rejects, Kant calls a theory of the supernatural 'hyperphysics', as opposed to the 'hypophysical' – literally 'sub-natural' – theory of scholastic hidden or 'occult' qualities.[17]

Fn. IV 411 If Kant is referring to the only extant letter (X 111–13, No. 62 [58]) by the philosopher and educationalist Johann Georg Sulzer (1720–79), we are forced to observe, first, that his reply is shamefully late: the letter is dated 8 December 1770, and by the time the

16 For a systematic classification of 'practical material' – rather than formal – 'grounds of determination', see *Critique of Practical Reason*, V 40; see also V 64 and Kant's classification of heteronomous principles of the will at IV 441–5 below. At IV 425.12–15 Kant warns his readers not to misinterpret the universal-law-of-nature formulation in the spirit of naturalism.

17 Kant criticises the assumption of 'occult' qualities as circular in R 3162, XVI 689: the thing to be explained is considered to be the cause. Meier defines them as qualities that are not cognised clearly and distinctly and assumed without sufficient reason (XVI 688–9).

Groundwork was published Sulzer had already been dead for more than five years. Secondly, Kant's correspondent writes that he is looking forward to reading a 'Metaphysics of Morals' soon. It took Kant about another decade and a half just to publish the present *Groundwork*. Thirdly, and perhaps most embarrassingly, Kant is replying to a question that his correspondent did not even ask. He discusses the thesis that, educationally, it is counter-productive to adduce many heterogeneous reasons in support of a single moral action. One of the results of Section I was that only volition on moral principles non-accidentally leads to action that coincides with moral commands. If, for example, we recommend a moral action to a child with reference to, over and above its true moral quality, a possible reward or some other desirable outcome, we are implicitly suggesting that there is less that speaks in favour of an equally obligatory action that is not accompanied by such pleasing effects. By contrast, if we praise solely the dignity of morally good volition the moral motive of reverence for the law is strengthened in comparison to other, potentially conflicting motives.[18]

¶ IV 411.8 This summary contains – besides numerous repetitions of the results established so far – the rather extraordinary idea that an investigation of pure practical reason can be even purer, in a manner of speaking, than a critique of pure theoretical reason.[19] This is due to the fact that practical principles, as completely formal, do not rely on the peculiarities of human cognitive capacities (e.g. space, time, the categories). It is only when practical principles are applied that anthropology – the study of human thought and agency – is needed. This points to a highly abstract conception of a metaphysics of morals as a theory of autonomy. It also foreshadows the rehabilitation of the traditional metaphysical objects of God, freedom and immortality in the *Critique of Practical Reason*.

¶ IV 412.15 We reached philosophical territory in the course of Section I. Kant now announces his intention of tidying up and advancing moral philosophy, which so far has been heavily relying on examples,

18 On the cognitive powers of pure practical reason see the chemical example of *Critique of Practical Reason*, V 92–3; on Kantian theory and practice of moral education the Methodologies of both the second *Critique* and the second part of the *Metaphysics of Morals*; see also what was said about the use of examples (IV 408.28).

19 But see *Critique of Pure Reason*, B 28–9 and A 14–15, which emphasises the essentially non-pure elements of moral philosophy.

by means of examining the 'practical faculty of reason', i.e. the will. Section II thus cleverly rejects popular moral philosophy by choosing the same starting point as Christian Wolff – an entirely *general* definition of the will – and then separating out pure and empirically conditioned volition – whereas Wolff did not realise this was necessary, if he had a clear idea of the task at all (see IV 390.27). At a more technical level, Kant now conducts an analysis similar to that of the previous section, leading to essentially the same result at IV 421.6–8 (cf. IV 402.8–9). Kant's chosen method also helps to identify the grain of truth within otherwise mistaken 'heteronomous' popular moral systems in the shape of the variant formulations of the categorical imperative.

2 *The doctrine of imperatives*

a. The will as the capacity to act in accordance with the representation of laws

¶ IV 412.26 Kant begins his discussion of the will, and the standards that govern it, with the rather unspectacular thesis that everything in nature obeys certain laws, i.e. the laws of nature, implicitly alluding to a second thesis that follows from the first: in so far as human beings are part of the natural world, they also work in accordance with natural laws. This second thesis is much more problematic than the first because it calls for a reconciliation of freedom and natural determinism.

However, man is not just a natural but also a rational creature – a necessary condition of possessing a will, defined as the capacity to act in accordance with, or as a result of, the representation of laws (*nach der Vorstellung der Gesetze*, IV 412.27);[20] which in turn is equated with acting 'in accordance with principles' (*nach Prinzipien*, IV 412.27–8).[21] The definition raises the question of what sort of principles or laws Kant has in mind at this stage, and how one is supposed to be able to act on them or in accordance with their representation. One idea might be that Kant is referring to maxims, i.e. *subjective* principles.[22] However, this interpretation must be ruled out, as the further course

20 'Idea' (Paton, Zweig) and 'conception' (Abbott, Beck, Ellington) are too pallid. It would seem that we must represent certain laws to ourselves to be able to act on them.

21 This distinction between wholly natural processes and action that consciously follows objective principles echoes the earlier distinction between action that merely coincides with duty and action from duty (see esp. IV 397.11–32). Action from duty presupposes a will that, like the human will, is capable of acting on moral principles.

22 This has been suggested by Paton, *The Categorical Imperative*, pp. 81–2. Cf. also R. Bittner, 'Handlungen und Wirkungen', in *Handlungstheorie und Transzendentalphilosophie*, ed. G. Prauss (Klostermann, 1986).

of the argument reveals. At this stage Kant is still speaking of 'the will' in general. The definition covers all possible kinds of will – a human will that is affected by sensibility, as well as a divine or perfect will. (Animals possess a faculty of choice – *Willkür*, *arbitrium* – but not a will in this sense.) A will cannot as such be the faculty or capacity to act in accordance with subjective principles or maxims because a holy will does not possess maxims. Nor does it act on interest or make use of incentives. A holy will is not subject to sensuous influences at all (see IV 413 fn.). It is, of course, determined by objective principles, laws in the strict sense of the word (see IV 414.1–5). If the above definition covers this kind of will it can only be this kind of law that Kant is referring to at this point: only rational beings possess a capacity to observe *objective* principles. They represent laws to themselves and then act accordingly. In both cases a will can effect that conformity with rational laws is non-accidental.

It is important to note that Kant calls the will a faculty or capacity (*Vermögen*). A capacity persists even if – as is the case with human beings when they fail to act rationally – it is temporarily left unused.[23] By contrast, a being endowed with a perfect or holy will always makes the best possible use of its volitional faculty. Kant calls the will 'practical reason' because it is the will's task to subsume individual cases under general principles or laws. These cases serve as a minor premise in a practical syllogism. The conclusion is the resulting action. In the practical as well as the theoretical sphere, drawing conclusions is the task of reason. In the *Critique of Pure Reason*, reason is generally defined as the 'faculty of principles' (A 299/B 356).

Kant then turns to the two kinds of will, first to the holy or perfect will that is 'infallibly' determined by reason. There is, in such a will, no competing force that might provoke it to stray from the right path. A being endowed with a holy will always elects to do what it recognises to be good through the laws of reason. The laws of reason are both objectively and subjectively necessary, and internal conflict is impossible. Things change dramatically when we turn to a will in which reason

23 The idea that reason is a capacity also explains the apparent contradiction that within a couple of lines at IV 412.26–31 Kant first identifies practical reason with the will and then speaks of reason as determining the will (see Paton, *The Categorical Imperative*, p. 80). The will is *practical* reason in that actions can be determined by rational standards. (*Thinking* about practical matters is not sufficient for 'practicality'.) When reason succeeds, it is practical in the literal sense; when – as is often the case with human action – reason fails, the elective part of the will or 'choice' (*Willkür*) is determined by non-rational forces instead. But failure was not inevitable, and the *capacity* persists.

does not 'solely by itself' invariably produce a corresponding action. This is the case with a will like the human will that is exposed to essentially subjective incentives rooted in the sensual side of human nature and potentially at odds with objective reason. It is only there that the distinction between subjective and objective principles makes any sense. For such a being, good action is contingent in the sense that *when* it occurs it was not inevitable because action from inclination was a genuine alternative. Such a creature is conscious of the fact that it does not automatically (subjectively) will what is (objectively) required. It is *necessitated* by the objective laws of reason.

b. Imperatives necessitate an imperfect will to act in accordance with laws

¶ IV 413.9 Introducing the notion of an imperative, Kant once again falls short of terminological precision. In the preceding paragraph, a 'will' was officially defined as the capacity to act in accordance with the representation of laws, i.e. in accordance with principles (IV 412.26–8); and these principles were shown to be objective principles. Now Kant somewhat tautologically defines a command as *the representation of an objective principle* (IV 413.9). This manner of speaking is, however, suggested by the idea that it is principles – as imperatives – that necessitate, not laws as such. To be a command, a law must not just be valid in the abstract; it must also have normative force for the agent. Consequently, the necessitating principle that corresponds to the representation of a law within a finite will is called a 'command (of reason)'; and the corresponding formula is its 'imperative'.[24]

¶ IV 413.12 Kant now provisionally connects the concept of practical value with the idea of an imperative: an action commanded by an imperative is in some sense good. He contrasts imperatives as the formulae of necessitating rational commands quite in general with the attractions of short-term pleasure, which tend to impede their realisation in action. The distinction between hypothetical and categorical imperatives follows two paragraphs further down.

Fn. IV 413–14 The note returns to the theme of possible different types of will, clarifying the notions of inclination (*Neigung*), need (*Bedürfnis*)

24 On imperatives in general see Kant's lectures on moral philosophy, *Collins*, XXVII 255–60, and *Mrongovius* II, XXIX 605–9. On the idea of an imperative as the *formula* of a command see the second *Critique*, V 8 fn.

and interest (*Interesse*). Inclinations are the 'source of need' (IV 428.14–15) because they are directed at some external object that the agent may or may not be able to effect: such an agent is not self-sufficient. It is only in a 'contingently determinable' will – a will that like the human will is susceptible to external, sensual influences – that principles of reason generate an interest, out of which an agent can then act, or fail to act. Any interest – as opposed to a brute desire – involves a rational contribution,[25] even though this may only reflect a partial judgement of reason: when we defy morality we may still act from interest, despite the fact that reason, all things considered, is altogether on the side of duty.[26] An agent is acting *from interest* if within such a will it is an inclination towards some object that is determining the rational laws of an action. For this, the object must seem 'agreeable' (*angenehm*, IV 413.38). We shall soon learn that this kind of action is action on hypothetical imperatives. As opposed to this 'pathological' kind of interest there is also the idea of a 'practical' interest, which occurs if a will *takes an interest in* the pure law of reason without being impelled by an inclination to do so. It is important to note that in the former case, as the present passage indicates, the will is interested in the realisation of the 'object of the action'; i.e. the state of the world at which inclination is directed. In the latter, practical case one is interested solely in the (moral) action itself. That is why when morality and inclination coincide, inclination will still be satisfied even if the agent acts solely from duty, as long as the action is successful. By contrast, reason will never be satisfied if an action that merely accords with duty is willed only indirectly, to satisfy some friendly inclination. In Section III, Kant discusses the uncertain possibility of an immediate practical interest in an action, which alone would make a categorical imperative possible (see IV 449.13–23 and IV 459.32–461.6).

¶ IV 414.1 The same laws can hold for perfect and imperfect beings, but they are descriptive for the former and prescriptive for the latter.[27]

25 Interest is defined as 'that which makes the will practical' at IV 459–60 fn. See also Kant's developed account in the second and third *Critiques*, V 79.19–24 and V 205–9.

26 See, again, the chemical example in the second *Critique*, V 93.5–6.

27 It is, however, difficult to see how the divine will could stand under laws that correspond to hypothetical imperatives, particularly of the pragmatic variety. As God lacks inclination, he also lacks the desire to satisfy the sum of all inclinations, i.e. to be happy. This difficulty may seem to be less problematic if, as perhaps suggested at IV 419.3–10, a pragmatic imperative is seen as the limiting case of a particularly intricate technical rule to realise an uncertain, unsteady and composite – if definitely given – end, rather than an imperative in its own right. Technical rules could more easily be seen to apply to the divine will because, presumably,

A perfectly good will is not even tempted to do anything other than what is good. Necessitation is absent; it does not face laws as commands. Imperatives, and consequently the opportunity to choose between good and bad actions, are expressions of human imperfection.

c. Imperatives, hypothetical and categorical: skill, prudence, morals

¶ IV 414.12 Kant now introduces the distinction between imperatives that command hypothetically and those that command categorically, between 'hypothetical' and 'categorical' imperatives for short. As the longhand expressions indicate, imperatives are distinguished by the way in which they command: some imperatives enjoin an agent to perform a certain action only on the condition that some incentive-based[28] end is given, others command independently of whatever – further[29] – purposes the agent may happen to have.

Kant has little interest in the grammatical or logical structure of linguistic expressions that correspond to imperatives.[30] In fact, it is crucial to note that Kant's classification of imperatives is not about the way we speak at all.[31] Categorical commands can be expressed in conditional sentences, and hypothetical imperatives in unconditional ones.

even God needs to realise his ends. Yet again, it is unclear whether he would make use of the same regularities that human beings follow when they act on technical imperatives. It is most likely that Kant first and foremost has the *moral* law in mind when he says that the same objective laws of the good hold for both kinds of will alike.

28 I have chosen this broad expression because hypothetical imperatives need not serve inclination, even if they always presuppose an end. In Section I, Kant often suggests that hypothetical imperatives depend on ends given by inclination because he wants to emphasise the formal and unconditional nature of the categorical imperative. It is immaterial to the argument that technical imperatives are also required for the realisation of moral ends, supported by the incentive of reverence, by making the use of a means indirectly a matter of duty (see what was said about 'indirect' duty, IV 399.3 above). 'Rules of skill' tell us how to realise *any* given end, irrespective of its source.

29 When Kant says that a categorical imperative represents an action as by itself objectively necessary, 'without reference to another end' (*ohne Beziehung auf einen andern Zweck*, IV 414.16–17; cf. IV 415.3), he indicates that moral action contains the end or purpose pursued within itself. A moral agent is directly interested in the obligatory action, and only secondarily and indirectly in its object; see IV 413 fn.

30 Similarly, the 'analyticity' and 'syntheticity' of imperatives introduced at IV 417.11 below is not identical with the analytical and synthetic judgements familiar from Kant's theoretical philosophy, but they are structurally analogous. See Introduction.

31 Philippa Foot falls prey to this confusion in her well-known article on 'Morality as a System of Hypothetical Imperatives', in *Virtues and Vices* (Blackwell, 1978). Rules of etiquette may appear to be categorical because we commonly expect everyone to comply, and rarely specify the conditions under which they are valid normative requirements. For the most part, however, they can undoubtedly be reduced to technical or prudential imperatives. Someone who has no wish to be accepted in polite society has no reason to adhere to its rules. Moreover, in some rare cases rules of etiquette may be indirectly supported by moral requirements, such as a duty not to offend one's hosts. Considered by themselves and in isolation, they completely lack normative force. The question whether such commands 'apply'

'If you wish to be a decent person, do not lie!' does not make veracity a conditional requirement because the protasis is immaterial. The policy of the prudent merchant not to overcharge children for the purpose of the long-term success of his business does not render honesty in financial matters merely conditionally good. Similarly, suppressing the condition required to make a mere regularity a normative command does not turn hypothetical imperatives into categorical ones. 'Play the piano!' is never required categorically and by itself, no matter how authoritative the person who makes the pronouncement might be. We need to be given a further reason to see whether it is justified. It would be justified, for instance, if you were justifiably committed to the end of becoming a concert pianist, or if you had a very good reason to annoy your neighbours. In fact, any kind of imperative, whether hypothetical or categorical, is by definition the formula of a valid command, i.e. it tells the person to which it is addressed to 'Do X!'. Kant is merely trying to point out that some commands are subject to motivational conditions whereas, surprisingly and problematically, others are not.

The present passage alludes to the tripartite classification of imperatives with reference to a contingent, actual or necessary end (see IV 414.32–415.5).[32]

¶ IV 414.18 Kant explains that all imperatives command the action in question as good and thus make it necessary for a will in so far as it is guided by reason. But that does not mean that they all command without conditions. A hypothetical imperative specifies the means towards an end that is presupposed. It commands actions that are instrumentally good. Such action is conditionally necessary. A categorical imperative commands action that is good and an end just by itself. Moral action is absolutely necessary.

¶ IV 414.26 A finite will like that of human beings is limited in at least two ways. First, it does not always know which of the various

is misleading. Hypothetical imperatives universally *apply* to everyone in the sense that the conditional is valid (e.g. 'If you desire to become a good pianist, practise regularly' is a valid rule in the way in which 'If you desire to become a good pianist, take hard drugs' is not); but they have *normative force* as imperatives proper – they translate into a command to 'Do X!' – only if the motivational condition is met and we have reason to realise the end in question. By contrast, the categorical imperative is a valid normative command because we – without any reference to our natural ends – impose the moral law upon ourselves.

32 On the distinction between hypothetical and categorical commands see *Critique of Practical Reason*, V 20.14–21.11, where all hypothetical imperatives appear to be technical (though not 'problematic'). No explicit provision is made for 'pragmatic' imperatives (see Kant's uneasy account of their possibility at IV 417–19 below).

options available is good. This epistemic problem occurs only in the case of hypothetical, i.e. non-moral imperatives. The agent must understand the workings of the world to realise a given end.[33] The categorical imperative, by contrast, is a (synthetic) principle a priori that does not depend on empirical information, which explains, at least in part, the alleged ease and accuracy with which in moral matters human beings are supposed to reach the right conclusion.

Secondly, a will like ours faces the problem that its maxims and consequently its acts of volition do not always coincide with rational commands. We all too often act contrary to the commands of reason. This danger is particularly severe with regard to moral imperatives because they are not supported by inclination.

¶ IV 414.32 Kant now expands the twofold distinction of imperatives to accommodate two different kinds of hypothetical imperatives, depending on whether an action contributes to a purpose that human beings contingently or non-contingently pursue. By contrast, the categorical imperative does not refer to the agent's purposes at all. Kant describes the purposes involved in these three kinds of imperatives with reference to the categories of modality (possible purpose – real purpose – necessary without a further purpose), and consequently names them 'problematic', 'assertoric' and 'apodictic', in accordance with the corresponding modes of judgement (see A 80/B 106 and A 70/B 95 respectively). These attributes refer to the manner in which the three kinds of imperative command. By contrast, the triad 'technical – pragmatic – moral' defines their domain: skill (*Geschicklichkeit*), prudence (*Klugheit*) and morality (*Sittlichkeit*).

This classificatory scheme is rather too tidy to be correct, as Kant later came to realise. Kant revokes the expression 'problematic imperative' in the unpublished 'First Introduction' to the *Critique of Judgement* as self-contradictory (XX 200.11–17). Such imperatives are to be known as 'technical' (belonging to 'art' or 'skill': Greek τέχνη, German *Kunst*), a term used alongside 'problematic' in the *Groundwork* at IV 416.29. The reason is that it is not imperatives themselves that are 'problematic' (depending on a specific problem, contingent) when they apply but rather the ends they presuppose. Technical rules become imperatives only when an agent desires a specific, individual, contingent end; by

33 See the illustration of the categorical imperative in Section I, IV 402.16–403.33. In the case of pragmatic imperatives a human agent encounters the additional difficulty that he or she must also get a clear picture of the balance of his/her divers ends; see IV 418.1–419.2.

contrast, the pragmatic imperative relies on an end that every human being pursues: happiness.[34]

¶ IV 415.6 'Problematic' imperatives of skill. The many technical imperatives serve any possible purpose a person may have. Judged just by themselves they are therefore 'value-neutral'. They do not raise the question whether the end pursued is itself morally (or even prudentially) good, i.e. whether it deserves to be pursued – which is why Kant argues that the teaching of skills is often overrated in the education of children. We learn more in the opening paragraph of the *Critique of Practical Reason*. Kant says that subjective principles (maxims) comprise under them several 'practical rules' (*praktische Regeln*), a term used for technical imperatives in the context of the second *Critique* (see *Critique of Practical Reason*, V 19.8). An agent draws on such rules as convenient. As opposed to his maxims, they are not an expression of his character. They are confined to the choice of means and do not concern the adoption of ends.[35]

This raises another philosophical problem about technical imperatives. It is important to put the instrumental goodness of essentially value-neutral technical imperatives in perspective. The fact that these rules are 'good for something' in a narrow, instrumental sense does not entail that, given a specific end the agent wants to pursue, they translate into fully fledged commands of practical reason. Indeed, Kant often associates technical imperatives with theoretical reason because of the way they track causal connections.[36] Thus even though a rule of skill may teach me how best to poison my rich uncle, there is no rational, all-things-considered command to poison him in such-and-such a manner – even if I do in fact intend to kill him because the end pursued is morally, and therefore rationally, illegitimate. That is one reason why at IV 416.20–5 below Kant indicates that only moral imperatives strictly deserve to be called practical 'laws'.

34 The term 'hypothetical imperative' may similarly seem to conflict with the idea that imperatives are *commands*, in this case commands of instrumental reason. It appears to suggest an element of indeterminacy or choice. The expression has led to the misunderstanding that a hypothetical imperative is a rational principle to the effect that one must either take the means or give up the end. 'Conditional' would have been a less misleading translation than the literal 'hypothetical'. *Hypothetisch* is the natural counterpart to *kategorisch*, i.e. absolute, unqualified, unconditional.

35 The morally relevant description of an action must always refer to an agent's maxim, not to a mere rule. It is *by means of* technical rules that one does what one wants to do.

36 See e.g. *Critique of Practical Reason* V 25.38–26.2, *Critique of Judgement* V 173.11–17, where technical rules are excluded from the realm of the practical altogether.

Kant could have avoided these problems if he had distinguished between 'technical rules' – value-neutral law-like regularities that may or may not relate to anyone's proposed actions – and 'technical imperatives', which command only (i) if an agent is in fact committed to realising a specific end and (ii) this end is morally – and perhaps prudentially – legitimate.[37]

¶ IV 415.28 Assertoric imperatives of prudence. The thesis that 'according to some natural necessity' (*nach einer Naturnotwendigkeit*)[38] all finite beings want to be happy follows from Kant's formal definition of happiness as the satisfaction of the sum of an agent's inclinations. Human beings are unavoidably equipped with natural desires. They can, in the long run, shape and cultivate them; inclination must take second place when it conflicts with morality; but all creatures like us still want them to be fulfilled, which will make them happy. By the standards of instrumental rationality, they are even rationally committed to realising their natural ends. The term 'natural necessity' may thus also imply that the desire to be happy is a necessary part of a human nature that comprises both rational and sensuous elements.[39]

A person who succeeds in consciously maximising his or her happiness is prudent. However, the assertoric imperative is still 'hypothetical' in the required sense. It is conditional on the agent's particular inclinations.[40] While all human beings share the 'formal' desire that their

37 Such a distinction may be intended at V 26.1–6 in the *Critique of Practical Reason*. It would correspond to the similar distinction between the moral law in the abstract and the categorical imperative, i.e. the moral law as addressing a finite will. Moreover, both technical rules and the moral law would be causal laws of their respective realms, nature and freedom.

38 Kant uses the indefinite article (rendered 'some' above) for two reasons. First, he does not wish to appear to endorse psychological egoism. For even though all beings like us possess the end to be happy, this end does not automatically translate into action. When the pursuit of happiness and morality conflict, we are free to be moral even at the expense of our personal well-being. The ends of happiness compete with those of morality, and are therefore hypothetical in the additional sense that they must make way if there is a conflicting categorical imperative. Secondly, as assertoric imperatives have a role to play in making people happy, Kant cannot say that happiness, as the satisfaction of all our inclinations, is a matter of natural causality in the standard sense, even if inclinations themselves are natural. On the necessity of our pursuit of happiness see also *Critique of Practical Reason*, V 25.12–14.

39 In the second *Critique*, Kant criticises the Stoic equation of moral virtue and happiness as ignoring the sensuous side of human nature (V 126.35–127.16). Somewhat ironically, the passage is highly reminiscent of later criticisms of his own theory as inhumane and overly demanding. As these passages show, Kant did not wish to dismiss the claims of our sensuous nature. He merely tried to put them in their place.

40 As well as, strictly speaking, the moral permissibility of pursuing the end in a given situation. Again, one might want to distinguish prima facie and all-things-considered rational commands. Like a technical imperative, an assertoric imperative is a rational command only on condition that the action recommended is morally permissible.

inclinations be satisfied, they possess different inclinations and thus differ with regard to their individual conceptions of happiness. This is why there can be no single, universal prudential imperative but at the very most more or less general prudential guidelines.[41]

Fn. IV 416 Kant argues that cleverness regarding one's dealings with others should not be called 'prudence' in cases in which one fails to use it to further one's own ends, which is the primary meaning of the word. It might thus be preferable to speak of 'worldly skill' (*Weltgeschicklichkeit*) rather than 'worldly prudence' (*Weltklugheit*). On the human inclination and capacity to influence other people see *Anthropology*, VII 271–2.

¶ IV 416.7 The apodictic imperative of morality. This kind of imperative does not presuppose practical ends. It selects and commands them. A categorical imperative immediately concerns just the action itself, and only secondarily and indirectly that which is to be effected by a moral action. Someone who wills the morally right action as a matter of principle has the right moral attitude or character, i.e. morally good maxims. Owing to its formal nature there must be a single, universal categorical imperative from which all particular (token) moral commands follow.

¶ IV 416.15 The three different types of imperative differ with regard to their psychological and normative status. Only the categorical imperative of morality concerns all human beings as rational creatures alike, which is why Kant would like to reserve the honorific title of a 'command' or 'law' for this imperative. By contrast, pragmatic imperatives are merely 'counsels' (*Ratschläge*) of prudence because – the general desire to be happy notwithstanding – what makes different people happy varies considerably. Such imperatives merely advise, they cannot count on universal acclaim. In fact, such 'counsels' hardly deserve to be called imperatives at all. Finally, technical imperatives are entirely dependent on an infinite variety of ends. These rules are universally *valid* in the sense that, to use Kant's own example, anyone who wishes to prepare a certain poison must follow the same pharmacological procedure; they *concern* only those who in fact intend to use this kind of poison for their purposes; and, as we saw above, they deserve to

41 Even they are severely affected by prognostic uncertainties about one's present and future ends. Kant returns to the complications of prudential imperatives and their 'possibility' at IV 417.27–419.11 below.

be called all-things-considered rational *commands* only if reason pronounces the purpose in question worthy to be pursued.

Fn. IV 417 Kant explains his choice of the word 'pragmatic' and, in accordance with the usage of his time, connects it with that which promotes human well-being; see *Critique of Pure Reason* A 806/B 834. At A 800/B 828 we learn that the ends pragmatic laws help us to realise are those 'recommended' – though not necessitated – by the senses. A 'pragmatic sanction' was an exceptional imperial or royal ordinance regarding the affairs of a community. The expression 'pragmatic history' was coined by Polybius (1.2.8); its original meaning is much disputed. Kant himself published an *Anthropology from a Pragmatic Point of View* in 1798.

d. How are all of these imperatives possible?

¶ IV 417.3 Kant now for the first time turns to the question of how these imperatives are 'possible', i.e. how it is possible for human beings successfully to be *motivated* to act on them as rational commands of their will, irrespective of whether or not the resulting action is graced with success. The possibility of a law's determining the will is a condition of its normativity.

Technical imperatives present the least difficulty because they are rules to which someone has reason to resort only if he or she desires to realise the end in question. Regarding Kant's thesis that whoever wills the end rationally also wills the necessary means, three things should be noted.[42] First, this is not the formula of a generic 'Hypothetical Imperative' but rather the answer to the very specific question of how imperatives of skill are possible, i.e. how a person can act on them. There is not much of a problem, Kant argues. My grinding coffee beans, for example, can easily be explained by my desire to make, and then to drink, a cup of coffee, in conjunction with a 'rule of skill'. Making coffee simply consists in – inter alia – grinding the beans. If I intend to realise an end and know the rule that helps me to do so, I act on my intention by taking the means, so far as reason is 'the decisive influence' on

42 Contra Thomas Hill, 'The Hypothetical Imperative', in *Dignity and Practical Reason in Kant's Moral Theory* (Cornell University Press, 1992); see also Christine Korsgaard's 'The Normativity of Instrumental Reason', in *Ethics and Practical Reason*, ed. G. Cullity and B. Gaut (Clarendon Press, 1997) and Bernd Ludwig's 'Warum es keine "hypothetischen Imperative" gibt', in *Aufklärung und Interpretation*, ed. H. F. Klemme, B. Ludwig, M. Plauen and W. Stark (Königshausen & Neumann, 1999). For the alternative view see also Mark Schroeder, 'The Hypothetical Imperative?', *Australian Journal of Philosophy* 83 (2005), 357–72.

one's action. One violates an imperative of skill if one prefers present pleasure to the realisation of an end to which one is committed.[43] (The answer is much more complicated in the case of pragmatic hypothetical imperatives, as the subsequent paragraph reveals.)

Secondly, this imperative does not command either to take the means or to give up the end. A hypothetical imperative in general commands nothing (IV 420.24–6). A fortiori, it does not command that we either take the means or give up the end. It is not the purpose of a hypothetical or any other kind of imperative to 'leave us options'.[44] Rather, it commands to take the means *when* the end in question is to be realised. It is doubtful whether Kant thought one could abandon an end proposed by nature – let alone reason – at will. Ends can only be rejected by categorical imperatives. Whether prudential guidelines can make us ignore some end proposed by inclination altogether is doubtful.

Thirdly, technical imperatives frequently make use of the results of synthetic knowledge. This does not affect their status as analytic *practical* propositions. According to Kant, mathematical instructions about how to halve an angle are synthetic epistemically, as is the poisoner's recipe, which draws on knowledge of the natural world and its laws. Analytic practical propositions are those according to which – in a judgement about what one ought to do – the volition is already contained in the agent's end, i.e. directly relates to his plans and inclinations (see IV 420 fn. below).

¶ IV 417.27 Pragmatic imperatives are more complicated than technical ones, and not merely because human beings often lack a clear conception of the ends they want to realise, i.e. what makes them happy overall. The empirical origin of our desires introduces an element of unpredictability – a pragmatic imperative is a piece of general advice (*consilium*) rather than a strict command (*praeceptum*). Even if at any given time there were, from the point of view of an ideal, omniscient observer, a single 'pragmatic imperative' tailored to the inclinations of an individual agent, he or she would not be in a position to apprehend it.

Furthermore, compared to technical imperatives, the question of what motivates us to comply with prudential guidelines when they

43 However, one might re-describe this case as a change of priorities: now one end (immediate pleasure) is preferred at the expense of another. It seems that we now need to call on higher-order considerations – such as prudence or morality – to decide whether the action is rational or not.

44 Hill, 'Editorial Material', p. 50.

conflict with the pleasures of the moment is much more urgent. Is there a special general desire to be happy overall and in the long run, over and above the individual inclinations that at present make themselves felt? In R 1028 (1776–8) Kant calls for a motive proper to prudence, in analogy with the moral motive of reverence. He makes it quite clear that not even cognitive perfection or omniscience is sufficient to produce prudent behaviour (XV 460.31–461.6). Such an incentive would make the pragmatic imperative a synthetic practical principle. It is, however, difficult to see how reason could generate an incentive not just for morality – which at least is its own product – but for something that is rooted in the alien realm of happiness and inclination (see IV 418.36).[45] Alternatively, if prudence does not possess an indigenous rational motive, could this gap be filled by reverence for the moral law? As we saw at IV 399 above, furthering one's happiness can be 'indirectly' a matter of morality; and the person who suffers from gout does, in the end, decide to follow prudential commands from duty. It seems unlikely, however, that duty would recommend that we further our happiness in all cases in which there is no direct duty to take care of. Duty's motivational 'backup function' would appear to be restricted to severe cases of misery, which raise moral problems of their own.

Kant's theory of hypothetical imperatives seems deficient, but we must always bear in mind that it is not the purpose of a *Groundwork of a Metaphysics of Morals* to develop such a theory. Rather, Kant needs the notion of a *hypothetical* imperative solely to clarify his novel conception of an unconditional *categorical* command of pure practical reason. Kant's interest in hypothetical imperatives is limited to showing what morality is not, just as in Section I the nature of moral action was revealed on the background of 'subjective limitations and hindrances' (IV 397.7–8). Readers will be disappointed if they turn to the *Groundwork* with the expectation of finding either an elaborate account of motivation or a complete theory of instrumental reason.

¶ IV 419.12 The possibility of moral action independent of any inclination is a particularly intricate matter. A categorical imperative is unconditional and cannot, like analytic practical principles, refer to ends an agent naturally intends to pursue. What it commands – an act of

45 The related question of whether the imperative of prudence is really an analytic practical proposition does not receive a satisfactory answer either, in the present passage or elsewhere, despite the fact that the question of how such an imperative is possible is supposed to be resolved with reference to its analyticity.

volition, solely for its own sake – is not previously contained in the agent's will. Relatedly, a moral action is also a free action and as such does not depend on a cause that precedes it in time. As the absence cannot be proven empirically – as far as experience goes, all actions *are* caused by previous events – sceptics contend that moral action motivated solely by reason and its own law is impossible (see IV 406.14–25); in other words, that it is never just the action an agent is rationally interested in but always its effect, conditioned by inclination. Kant is fully aware of the problematic nature of the assumption implicit in Sections I and II that such moral action is possible.

¶ IV 419.36 The fact that the possibility of moral action cannot be decided empirically calls for an a priori investigation of the question of how a categorical imperative is possible, to be postponed until Section III. Kant repeats his thesis that only the categorical imperative is in the strict sense a *law* for the human will (see IV 416.20–3). It is not clear that the possibility of 'pragmatic' hypothetical imperatives has been explained, but it may for the moment suffice to say that the categorical imperative is clearly problematic in a way a hypothetical imperative is not.

¶ IV 420.12 Like everything that is great and good in Kant's philosophy – the propositions of arithmetic and geometry; in particular, the categories – the categorical imperative is a synthetic principle a priori. It needs to be substantiated by means of a 'deduction' in Section III.

Fn. IV 420 Kant says that a *practical* synthetic proposition a priori is characterised by the fact that the act of volition commanded is not entailed by the ends the agent intends to pursue. There is therefore a structural analogy with the synthetic principles in the theoretical sphere. A theoretical synthetic proposition provides us – or at least claims to provide us – with new *information* not previously contained in the concept of some object of cognition. Similarly, the categorical imperative *commands* something new, a practical end not previously contained in the will. The synthetic character of the moral law depends on the specific nature of the human will. Such a practical law applies 'analytically' within a rational will that is not subject to the temptations of inclination, such as the will of God. For him, morality is not prescriptive. It follows from his 'natural' intentions.

3 The categorical imperative

a. Derivation of the general formula of the categorical imperative from its concept

¶ IV 420.18 Kant explicitly announces his intention to postpone the question of the possibility of an unconditional imperative until the final section of the *Groundwork*. Yet even if the possibility of agents' *conforming* to such an exacting standard cannot be derived from its concept, we might still be able to reveal something about the *content* of an absolute and formal practical law – a law that presupposes no specific end – i.e. about what this law commands.

¶ IV 420.24 Hypothetical imperatives are conditional. To enter into force as commands of reason they depend on the end an agent intends to pursue, supported by some incentive. That is why a hypothetical imperative as such commands nothing at all. The agent needs to wait until a specific end is given. Only then is he in a position to start to determine on which causal regularities to act in the situation in question, i.e. what he *ought* to do. By contrast, Kant argues, that which an unconditional and universal categorical imperative commands is entailed by its very concept. It is not any specific law that determines the will when it obeys a categorical imperative, but rather the universality of a law *as such*, which happens if and only if the maxim of an action can consistently be willed to be valid universally.[46]

This derivation of the formula of the categorical imperative has provoked much criticism.[47] Commentators often diagnose an unbridgeable 'gap' at this point of the argument:[48] Kant rather too hastily, they argue, moves from the universality of the law of one's volition to the general formulation of the moral imperative; and admittedly it is not

46 Yet even in the case of the categorical imperative the concept yields not individual token commands but rather the law that allows us to select concrete content in specific situations. (The fact remains that it is only in the case of the categorical imperative that a formal restriction leads to a material command.)

47 As well as the parallel move in Section I, see IV 402.4–9; see also *Critique of Practical Reason*, V 29.14–22.

48 See Paton, *The Categorical Imperative*, p. 72, B. Aune, *Kant's Theory of Morals* (Princeton University Press, 1979), pp. 29–30, and Wood, *Kant's Ethical Thought*, pp. 78–82. Essentially the same question was already raised by G. A. Tittel in 1786, who asks why volition based on, for example, an interest in universal happiness should be unfit for universal legislation (*Über Herrn Kant's Moralreform*, p. 51; cf. p. 54). For a recent comprehensive discussion see S. Kerstein, *Kant's Search for the Supreme Principle of Morality* (Cambridge University Press, 2002), esp. pp. 73–94.

entirely obvious that the universal law that appears to human beings as an unconditional moral imperative can only take the shape of the thought experiment described above. Might one not, for instance, make the principle of impartially maximising utility a universal law, and act for its sake, from moral conviction?[49] Several things should be noted. First, even if the principle of utility were compatible with Section II so far, it certainly cannot be derived from Kant's argument at this point, whereas there have been indications of what a categorical imperative might look like all along. The introduction of the new candidate therefore seems rather ad hoc. Secondly, it is difficult to see how a rational creature could will *a law that there be* universal happiness, rather than even universal happiness itself, for its own sake. In utilitarianism, it still has to be the happy effect that is willed qua good, rather than action on a certain principle – could a thoroughgoing consequentialist really applaud actions for the sake of a utilitarian law if they did not realise the (indirectly) intended effect? But if the effect is directly intended, the law an agent follows is instrumental, not a law that is willed for its own sake. At this point, the agent cannot will a moral action *because* it is good – no value is presupposed. It must be willed because it is commanded by the moral law. It is that act of volition that subsequently possesses moral value. Willing a *consequentialist* principle like the principle of utility *for its own sake* would be a particularly schizophrenic case of 'rule worship'. Thirdly, and relatedly, the value of the principle of utility would have to be derived from the universal desirability of happiness, i.e. some positive value. It does not meet rational approval in itself – at the very least if we accept Kant's notion of (pure practical) reason, which is the only one currently on offer. The senses like happiness; but reason does not revere it.

49 The other 'universal law' that is often seen as a rival of the categorical imperative at this point is the principle of the egoist who grants that others may be equally selfish. However, it would seem that the egoist's challenge is not relevant in the present context. An egoist, almost by definition, acts from inclination, i.e. for the sake of some (selfish) effect to be brought about in his favour, *not* for the sake of the volitional act itself; and it has already been established that this must be the case with moral volition. The parallel discussion in Section I (IV 397–401) provides us with ample resources to dismiss an egoist's claim to moral permissibility at the present stage of the argument. Mere coincidence with some universal law is not enough. (An egoist acting for the sake of – rather than in mere conformity with – the *principle* of selfishness would be a philosophical curiosity and, arguably, not even selfish in any recognisable sense at all.) However, even if we accept Kant's identification of universal regularity *überhaupt* and the 'formal' formulation of the categorical imperative, the question whether Kant's ethics can successfully reject selfish principles has not yet been answered. What remains to be shown is that these principles fail to pass the test of Kant's universalisation procedure. See the examples at IV 421–4 below.

To bridge the gap, we should first note that Kant is no longer interested in specific laws of the will – which must be universal, if restricted by specific ends and therefore not formal – but rather regularity, 'conformity with law', as such. Kant speaks of 'the law' in the singular throughout this paragraph. We are not moving from singularity to universality, but rather from one level of universality to another: from conditioned to unconditioned universality.[50] Kant's thought is simply that if there is no *specific* (material) law, and there must be a law that determines the will in moral action, it – its maxim[51] – must conform to formal[52] law *überhaupt*. But what does that mean?

As the footnote attached to this paragraph indicates, Kant is assuming that in acting on imperatives we actively conform subjective standards to objective ones. When we act in accordance with hypothetical imperatives, we conform the law of our volition to the empirical regularities we follow for the sake of the intended effect. Such action is universalisable, if only in a limited way: I must be able to act on a *rule* that has normative force for everyone who pursues the end in question. However, the universality of a categorical imperative is, by definition, unrestricted; and I must ask myself whether the principle implicit in what I am about to do, my *maxim*, can conform to universality as such; i.e. whether it can be adopted not just by those who share some interest in an end to be promoted or brought about, but by *all* rational agents like myself.[53] If we accept the basic elements of Kant's moral psychology, there is no gap. Of course, towards which actions a standard of conformity to universal law *überhaupt* leads us – indeed whether it can be action-guiding at all – remains to be seen.

50 The fact that he is looking for some law other than instrumental regularity to determine the will is explained by the fact that the will, by virtue of being a causal power, must always conform to some law to be active (see IV 446.15–21). The will was defined as the capacity to act in accordance with laws at IV 412.27–8 above. Imperatives, as rational principles of the will, must therefore contain some law or other.

51 Kant now explicitly relates universal regularity as such to the *maxim* of an action, not to the action itself, which was merely implicit at IV 402.6 in Section I above. That which determines what we do, and which a relevant description of the action must refer to, is – in moral cases – the maxim, in instrumental cases a rule. These are subjective analogues of objective standards; ultimately, no act of volition is without its maxim.

52 The categorical imperative is formal not in the sense that it can be derived from its mere concept, but because it presupposes no end that we intend to achieve. Rather, it determines the end that we ought to incorporate into the maxim we act on. As we need no matter to begin with, we can derive content from a mere law, i.e. the concept of an *imperative* as such.

53 Instrumental rules cannot be informed by universality as such because universality alone does not help us to achieve an intended purpose. Maxims cannot (just) conform to specific universal regularities because that would leave them without an ultimate, non-consequentialist justification. We cannot use instrumental laws to evaluate our fundamental ends.

Fn. IV 420–1 Kant re-states his official distinction between subjective principles (*maxims*) and objective principles (for a will that acts on maxims: *imperatives*) in more technical terms, as is fit for the purpose of Section II (cf. IV 400 fn.). Maxims can and ought to conform to imperatives, but this requires an act of free choice. They are the principles on which a person *does in fact act*. Maxims refer to the agent in the first person singular, and specifying the end of action they are expressive of the values to which a person is committed. It is important to note that the categorical imperative concerns maxims: an agent's fundamental attitudes that determine what he does, not single external acts. The moral quality of the latter is neutral.

b. The general formulation

¶ IV 421.6 We have now, by a somewhat more technical route, reached the point which marked the end of Section I: in a given situation, we are permitted to act on a subjective principle that can serve to determine the causal powers of our will *only* if we can consistently will that, by virtue of this very action,[54] we effect that it is universally adopted and acted upon. Kant calls this *the* categorical imperative.

As the word 'only' reveals, this imperative is first and foremost a criterion of moral permissibility. We must again consult Kant's views on moral psychology to understand its workings. As indicated at IV 397.11–21 above, he assumes that practical options are limited at any given time. All possible action happens for the sake of an end, supported by an incentive. We can act for the sake of immediate pleasure, of some higher-order end (ultimately happiness) or, in moral action, for the sake of the action itself. It is our maxim that determines our actual choice.[55] If in a relevant situation an available maxim passes this test we *may*, not *must*, act on it. As Kant's illustrations indicate, a specific duty arises when in a specific case only one maxim is both relevant and morally possible (permissible).

Kant makes it very clear that the present paragraph contains the one canonical and general formulation of the categorical imperative. 'There is only a single categorical imperative', he says, 'and it is this: . . .'. Other apparently unconditional imperatives are either

54 That appears to be the meaning of the awkward expression 'through which you can at the same time will' (*durch die du zugleich wollen kannst*) at IV 421.7–8. Cf. the formulation of the categorical imperative in the *Critique of Practical Reason*, V 30.38–9. Cf. also *Metaphysics of Morals*, VI 255.6–13.

55 It does not command that we act on maxims; it assumes that we act on maxims and then commands that any maxim we act on conforms to universal law as such.

variations of this principle or individual 'categorical imperatives', i.e. particular applications of this principle, like an unconditional command not to lie.[56] The unique status of the general formulation is confirmed when Kant summarises the three variants at IV 436.8–10 below. The law-of-nature formula with its four illustrations, which immediately follows the present formulation, is the first of three variations. The fact that there is only one supreme law is again a consequence of its nature as an unrestricted practical law.[57]

¶ IV 421.9 Once again Kant points out that at present he is concerned only with what follows from the common concept of duty. It is still possible that the concept of duty is 'empty', that it does not correspond to anything real, i.e. that the categorical imperative does not apply to human beings. Yet even now that we have been presented with the official formulation of the categorical imperative, one part of the analytic project remains. Kant must make the connection between the new principle and commonly acknowledged token obligations. Accordingly, the illustrations that follow the variant formulations of the categorical imperative demonstrate how different kinds of duty are entailed by the formula of the categorical imperative. Kant sets out not to discover new duties hitherto unknown but to derive various known duties from a single new principle.

4 The first variant: universal laws of nature

a. The universal-law-of-nature formulation

¶ IV 421.14 Kant now makes the transition from the general formula to the first of three variations:[58] the so-called 'law-of-nature formulation'. We progress beyond Section I in substance, not just in style. In the

56 It is curious that Kant chooses to give examples for the three variations but not the general formula. There are several, potentially complementary, reasons for this: first, Kant thought the application of the general formulation had already been dealt with in Section I (IV 402.16–403.17) and, secondly, he optimistically thought that the general formulation could be easily applied by everyone and does not stand in need of elaborate illustration (see IV 403.18–33). Thirdly, the variations of the present section fulfil the dual purpose of winning over philosophical rivals and of progressing towards the notion of autonomy in a novel moral metaphysics, which arguably makes the project of illustrating them rather more urgent.

57 It would, of course, have been somewhat unfortunate if the search had produced more than one 'highest' principle of morality, but now that it has been shown to be a purely formal law this danger has been averted. By contrast, there are substantially different technical imperatives, depending on the ends they presuppose; and prudential imperatives differ from person to person – no single prudential counsel is universally valid.

58 This formulation is a mere variant because, like the other two, it rests on an analogy: that of universal laws as laws of nature. See the overview at IV 436.15–18.

Critique of Pure Reason, nature in a *formal* sense is said to be 'the connection of determinations of a thing, in accordance with an inner principle of causality'.[59] The human will is a kind of causality because it effects changes in the visible world, in accordance with certain regularities. Viewed as natural events, human actions, permissible or impermissible, happen according to natural laws.[60]

To find out whether a possible maxim is morally legitimate or not, we must engage in the following thought experiment. Imagine a world in which the maxim on which we are about to act is not freely adopted, but rather part of the causal structure of the world. (Otherwise the creatures inhabiting this world are just like us.) The principle in question then naturally determines their actions in all relevant situations – past, present and future – in a manner analogous to the way in which instinct has the power to determine the behaviour of animals (see IV 423.13). Kant argues that, if the maxim we put to the test is immoral, such a world would somehow be flawed, incoherent, impossible. In our own world, of course, human beings face the choice between different maxims, which are then visibly expressed in observable actions, i.e. causal processes that obey natural laws. They are justified in acting on a maxim in question only if that action would still be feasible if by virtue of their causal act human beings *universally* behaved like that.[61] This idea will soon be illustrated by means of four examples.

Why does Kant – tentatively, cautiously[62] – offer this new formulation? Because, like the other two variants of the categorical imperative, it reflects the grain of truth contained in an otherwise mistaken popular ethical theory. The target of the first variant formulation is easily identified. It is the Stoic idea that a morally good life is a life in harmony with nature,[63] which was still popular with philosophers like Alexander Gottlieb Baumgarten[64] and Christian Garve, whose

59 A 418–19/B 446 fn.; see also the Typic of the second *Critique* and V 43.13–14.

60 Kant's exhortation to act as if one's maxim were to become 'a universal law of nature' is not a tautology.

61 Again the question arises how it is possible that all human actions are determined, as appearances, by such causal laws; in other words how moral freedom of the will and natural determinism can be reconciled.

62 At IV 421.17–18 the variant is introduced with the words that the universal imperative of duty 'might also be stated as follows' (*so könnte der allgemeine Imperativ der Pflicht auch so lauten*).

63 See Long and Sedley, *The Hellenistic Philosophers*, vol. I, pp. 394–401, vol. II, pp. 389–94, especially Stobaeus 2.75, 11–76, 8 (63B) and Diogenes Laertius 7.87–89 (63C).

64 See e.g. Baumgarten's *Initia*, §§ 45 and 46. The principle is criticised in the lectures on moral philosophy as at best a principle of prudence, not a moral principle (*Collins*, XXVII 266.10–19), i.e. according to the Kant of the lectures, the Stoics, like all moral philosophers before him, confuse pure and empirical practical reason.

annotated translation of Cicero's *De officiis*, published in 1783, is said to have inspired Kant to write the *Groundwork*. As early as 1770 Kant says that living in accordance with nature does not mean 'living in accordance with the impulses of nature but rather with the idea that is the foundation of nature' (R 6658, XIX 125). He goes on to say that nature and freedom are opposed to each other, and that the moral law is not a law of nature. Kant makes systematic use of this idea in his exposition of the present law-of-nature formulation of the categorical imperative.

As in the moral views commonly held by non-philosophers, universal pure practical reason obscurely makes itself felt in non-Kantian ethical theory, even if such a theory apparently commits the blunder of locating the principle of moral action in the workings of physical nature. Sophistication can corrupt morality; and moral philosophy is often more corrupt than common ethical thought (see IV 404.23–8) – which is precisely why the first two 'transitions' are necessary. The second transition thus serves a dual function. It represents not only Kant's own progression towards a moral metaphysics, but also the passage that popular moral philosophers have to make in order to come to accept Kant's theory.

b. Application of this formula to the four examples of duty

¶ IV 421.21 Kant calls his classification of duties as duties to the self and to others as well as perfect and imperfect duties 'common' (*gewöhnlich*); it is, however, unconventional in that it does not include duties to God. In the *Metaphysics of Morals*, Kant argues that we can only have duties towards similarly moral creatures with which we are acquainted (see VI 442.16–18). The rationale for the present classification is further discussed at IV 423.36–424.37.

Fn. IV 421 This footnote has caused several misunderstandings. Three comments. First, the classification is not 'arbitrary' (see translations by Abbott, Paton, White Beck, Ellington, Zweig), as the word *beliebig* suggests to the modern reader. Kant merely wishes to say that he has adopted this division because it suits his present purposes. 'Possible' (*möglich*) is given as a synonym of *beliebig* in the *Critique of Pure Reason* (A 74/B 100). The above classification of duties is legitimate, but rather coarse, as well as incomplete. For instance, it fails to specify sub-classes of duty within each of the four categories; it does not pay tribute to

the distinction between right and ethics that dominates the later *Metaphysics of Morals*; and it obscures the fact that violating a duty to others by, for example, making a lying promise also fundamentally violates a duty to the agent's own self (see Kant's essay on lying, VIII 426 fn.). This may serve to explain Kant's reservations at this point. Secondly, Kant gives the misleading impression that it might be permissible to make *exceptions* from wide duties in favour of one's inclinations.[65] It is probably with reference to this passage that in the *Metaphysics of Morals* he says that even wide duties must not be relaxed for the sake of inclination (VI 390.9–14). The good will cannot give way to any inclination – transgression of duty is defined precisely in terms of making an exception in favour of inclination at IV 424.19–20 below. What Kant intends to say is that perfect duties do not allow of exceptions for the sake of any substantive end that we wish to promote, moral or otherwise, whereas imperfect duties initially permit trade-offs between morally worthy ends (or 'grounds of obligation'; see *Metaphysics of Morals*, VI 224.9–26). Thirdly, Kant introduces perfect duties towards one's own person, in violation of a convention that goes back at least to Pufendorf's *De jure naturale* of 1672. Kant departs from tradition because, their character as perfect duties notwithstanding, perfect duties to the self cannot be externally enforced.[66]

¶ IV 421.24 The first illustration: suicide. The case of the man who desires to kill himself reveals that the categorical imperative goes well beyond universalisation in the standard modern sense of the term. The example does not turn on the question of what the world would look like if everyone shared the unfortunate man's attitude. 'Interpersonal' universalisation across agents yields duties to others. At present Kant is trying to establish a duty to the agent's self, which depends on the notion that it must be possible to will a maxim as an atemporal principle. A law lacks universality if it fails to apply in all relevantly similar circumstances, not just at present but also in the past and in the future. (A truly universal law cannot change.) That is why, at IV 424.19 below, Kant says that some actions violate the commands of duty because the

65 M. Gregor draws this conclusion in *Laws of Freedom* (Blackwell, 1963), p. 96; see also T. E. Hill, 'Imperfect Duty and Supererogation', in *Dignity and Practical Reason in Kant's Moral Theory* (Cornell University Press, 1992) p. 148, and 'Meeting Needs and Doing Favors', in *Human Welfare and Moral Worth. Kantian Perspectives* (Clarendon Press, 2002), p. 214.

66 On the division of duties see *Metaphysics of Morals*, VI 240 and VI 390–1.

agent takes the liberty of making an exception from a generally valid law in favour of inclination 'for just this once'.[67]

This first example purports to prove that it is immoral to commit suicide on essentially selfish grounds, for example because one is tired of life. The maxim rejected is one that rests on inclination. Kant does not wish to establish a prohibition of suicide as such. In the *Metaphysics of Morals* he intimates that in exceptional cases suicide might be legitimate or even obligatory (VI 423–4) – which raises problems of its own, particularly in the context of the second variant's injunction never to use oneself as a mere means. Applying the first variant he rather intends to show that a generally accepted moral duty not to throw away one's life can be derived from the newly discovered principle called the 'categorical imperative'.[68] The suicide's maxim to shorten his life when its longer duration promises to contain more pain than pleasure is motivated by inclination.[69] It is a 'principle of self-love' (IV 422.7), which cannot become a universal law of nature. It is commonly the task of pleasurable sensation to impel agents to further their life. Kant detects an inconsistency in an imagined natural order that now uses this very sensation to destroy it.

The word rendered 'task' above, *Bestimmung* (IV 422.9–10), might suggest that purposive nature plays an unduly prominent part in Kant's foundational story of morality. It is not at all clear that Kant ought to, or indeed can, avail himself of teleological premises at this point. However, there may yet be another possibility. The 'vocation' or 'task' of the striving for pleasure might not be set by nature but rather by the agent himself. Pleasure is normatively inconclusive, but the agent's attitude

67 There are indications of this throughout Kant's work. According to the formulation of the moral principle in the lectures, our actions must cohere (*übereinstimmen*) with a rule that is valid 'at all times and for everyone' (*Mrongovius*, XXVII 1427.1–4). (The passage is missing in *Collins*.) See also *Critique of Practical Reason*, V 36.15. The idea that all duties – and those to one's own self exclusively so – depend on whether they can be sustained across a whole life has been emphasised by O. Höffe, 'Kants nichtempirische Verallgemeinerung', in *Grundlegung zur Metaphysik der Sitten. Ein kooperativer Kommentar* (Klostermann, 1989), p. 221. For an extensive discussion see J. Glasgow, 'Expanding the Limits of Universalization: Kant's Duties and Kantian Moral Deliberation', *Canadian Journal of Philosophy* 33 (2003), 23–48, who also rightly notes that the categorical imperative always makes use of atemporal universality, even in cases of duties to others (p. 40).

68 He is currently 'enumerating' (*herzählen*) commonly recognised duties; see IV 421.21.

69 The actual formulation at IV 422.4–7, which seems to make self-love part of the maxim, is a good illustration of the fact that maxims do not arise in a motivational vacuum. They are always tied to specific incentives, which offer the general principles implicit in a corresponding action to reason for evaluation. An agent cannot act on a purely speculative maxim if he lacks the corresponding motivation. A fortiori, human beings cannot adjust their maxims to specific circumstances to make them 'universalisable'. Any such story that an agent uses to justify an action to himself is simply an act of self-deception.

towards it must be consistent. The idea would then be that someone who normally takes the striving for pleasure to be a reason to further his vital interests should now, all of a sudden, consider it a reason to destroy his life. This would, at least, be a recognisable contradiction. One must not expose one's life to the contingencies of a cost-benefit analysis. As Kant puts it in the second *Critique*, this would make a 'lasting order of nature' impossible (V 44.11).

¶ IV 422.15 The second illustration: false promises. Fortunately, this example is more obviously successful than the first. As in the previous case, the person who would like to make a false promise is prompted by inclination, and the corresponding maxim is again called a 'principle of self-love' (IV 422.24). Inclination makes itself felt first, but it does not of course automatically produce the action it favours. The agent is free to pause and ask himself whether the action proposed by inclination conforms with rational requirements, i.e. whether he can will the implicit maxim to be universally adopted. It is possible, but by no means certain, that a false promise will turn out to be beneficial, as the first discussion of this case in Section I reveals (IV 402.16–403.33); but the agent would certainly contradict himself in action because his maxim cannot subsist as a universal law of nature.

Why does Kant think there is a practical contradiction? Not simply because the universal validity of the maxim under review as a law of nature would render promises impossible, or because undermining this institution might have undesirable effects. Rather, making a false promise, the agent must implicitly rely on this institution to be able to achieve his end, i.e. to escape his financial difficulties. The agent does not merely intend to say something he believes to be false. He wants another person – someone who does not know the promise to be insincere – to trust him. He is trying to make an exception for himself from a principle he wants others to observe.[70] If his maxim were to govern human behaviour universally and naturally, in a quasi-mechanical fashion, he would not be in a position successfully even to *make* a promise because the other person would not believe him. The

70 This marks the difference between 'consequentialist universalisation' and Kantian morality. The mistaken allegation of Kant's implicit consequentialism goes back as far as G. A. Tittel's 1786 *Über Herrn Kant's Moralreform*. 'Does not', he asks, 'this very law point to the consequences, that such a maxim would have for myself and others, in accordance with its universality?' (p. 14); see also p. 33, where the law is said to be 'an empty and sterile concept' incapable of application if it does not refer to empirical 'consequences and effects', and Tittel's complaint that he cannot make sense of the law's 'empty form', p. 88.

other person would know from experience that promises are not to be trusted.[71] To put it somewhat differently: a world in which someone successfully makes a promise although promises are not accepted is *inconceivable*.

Two further notes: first, the agent cannot *will* the universality of his maxim either, because that would frustrate his end. If an agent cannot make a promise, a fortiori he cannot obtain a loan by means of a false promise. Maxims which fail the stronger criterion of a 'contradiction in conception' (one cannot conceive of a world in which the proposed maxim holds as a universal law) thus also fail by the standard of the weaker criterion, the 'contradiction in the will' (according to which one cannot will such a world); cf. IV 424.1–14. Secondly, note that just like the first this illustration relies on the timelessness of universal laws, even if it also makes use of the more familiar kind of interpersonal universalisation ranging over all agents. It is not sufficient that promises start to be undermined at the time of the agent's lying promise – he would not undermine his own action and no practical contradiction would arise. By contrast, if the maxim had been a timeless law of human nature, the institution of promising would not have evolved. Attempts to make false promises in times of need would be as predictable as the movement of my dangling leg when the doctor's hammer hits the right spot on my kneecap.

A good indication of someone's violating a strict duty towards others is the fact that the other person would not co-operate if he or she was aware of the agent's intentions. It is clear that the person in Kant's example must keep the true purpose of his promise to himself in order to succeed. If the other person was prepared to part with his money, even without the expectation of getting it back within a certain time, there would be no need to deceive: he could simply ask for the money.[72]

¶ IV 422.37 The third illustration: developing one's talents. Kant's attempt to castigate a maxim of laziness as immoral is the weakest of the four illustrations. This is not a coincidence. A look at the lectures on

71 Similarly, a thief is not simply undermining the institution of property. He wants to *keep* the stolen thing, and that is how the contradiction arises. We may be able to will a world without property, but we cannot will such a world and at the same time intend to possess property ourselves.

72 On the criterion of prescience and knowledge of intentions see R 6734, XIX 144: 'An action is wrong [*unrecht*] in so far as it is impossible if others presuppose these principles in us, e.g. lie. It is impossible to cheat on someone who knows that one wants to cheat, or breach of trust in contractual matters. It is also impossible to will and to condone such an action as a general licence.'

moral philosophy reveals that Kant had difficulties making up his mind whether developing one's talents was a matter of moral importance at all, and his reluctance is still apparent in the *Groundwork*.[73] The second edition of 1786 contains a rare emendation: Kant adds that the lazy person's capacities 'serve him <and are given him> for all sorts of purposes' (IV 423.15–16) – which raises the worry that Kant has to rely on teleological principles to make an allegedly purely formal ethical theory work.[74]

To make the present argument work, Kant thus has to steer clear of the twin dangers of relying on groundless teleological premises and substituting a prudential argument for a moral one. The latter danger is particularly apparent. He recommends a maxim of furthering and developing one's natural talents in the face of the temptations of short-term amusement because there are 'all sorts of purposes' that one may wish to pursue later in life. This makes self-development look like a matter of long-term prudential planning rather than a command of duty, but the two kinds of requirement may yet be distinct. Kantian prudence requires only that we cultivate those of our talents that serve our natural interests. After all, the pragmatic imperative has as its condition the sum of an agent's inclinations. There will be considerable uncertainty as to which of our individual talents we ought to cultivate, but the basic fact that prudential advice depends on the contingencies of human nature remains.

By contrast, a duty to develop one's talents must be unconditional, like every other moral command. Towards the end of the present passage Kant says that 'as a rational being' the agent 'necessarily wills that all the capacities in him be developed' (IV 423.13–16); and this is not because *all* the capacities may be needed in the future, but because given a choice of principle to develop one's talents or to neglect them only the former can consistently be willed. The contradiction incurred is similar to that in the fourth and last example: all agents are committed to realising their ends, and this is inconsistent with systematically neglecting that which they need to realise them on the basis of an ever-repeated momentary cost-benefit analysis. This argument is formal, and as in the first example the universalisation test applied to the maxim in question is temporal. Moreover, it does not rely on purposes assigned

73 In the lectures, Kant criticises the author of his ethics textbook (Baumgarten) for counting the perfection of one's talents amongst moral duties to the self. They are morally relevant only indirectly (see *Collins*, XXVII 363–4). He has changed his mind.

74 This worry features prominently in Paton's work, *The Categorical Imperative*, p. 17.

to human capacities in a grand teleological scheme. It quite suffices that the agent has some purposes of his or her own.

Even if the details remain obscure, the third example, like the subsequent fourth, demonstrates how positive duties can be derived from an essentially negative criterion. If a maxim cannot be sustained as a universal law, one must adopt the opposite attitude. But precisely because of this more indirect procedure the connection with action is more remote. Wide duties do not, like strict duties, directly command that we refrain from certain actions. They urge us to adopt a maxim to act in a certain manner when the occasion arises, i.e. they require judgement, as Kant puts it in the second *Critique*, as to whether something is a 'case of my maxim' (see V 27.26). Furthermore, maxims of wide obligation can on certain occasions (though not per se) be in conflict with each other and call for arbitration.

¶ IV 423.17 The fourth illustration: beneficence. Imagine someone – for example the indifferent but honest man we encountered in Section I (IV 398.27–36) – who by nature lacks any interest in the affairs of his fellow human beings. He has no desire to infringe on their rights, but he is not inclined to help anyone who needs assistance either. Admittedly, someone who has no immediate intention either to help or to receive help is not guilty of the gross inconsistency of those who expect their fellow human beings to help them but are not prepared to help others in return.[75] To establish a duty of care towards others, Kant argues that this 'way of thinking' (*Denkungsart*), if universally adopted, would lead to a contradiction in this man's volition.

The maxim of the indifferent person is far from innocent. Kant assumes, not implausibly, that human needs present claims that cannot simply be rejected. Hunger, for instance, cannot just be reasoned away. A hungry person desires to eat, and it is entirely rational for him or her to hope for the assistance of those who could easily help. This assumption goes some way towards explaining the impossibility of a principle not to care about the hardship of others. If we cannot but desire the help of those more fortunate than ourselves, it is inconsistent for us to deny our assistance to those who are now in that kind of predicament.

75 Would this rest on one maxim or two? Would it count as producing a 'contradiction of conception', as does the second example? Would it involve some self-deception? Or be a kind of moral schizophrenia? Kant does not expand. Nor does he say how such a maxim of selfishness, rather than indifference, if rejected, would lead to the establishment of a duty of beneficence.

However, the contradiction in willing – though not in thinking – that results when this maxim is conceived as a universal law of nature is also due to the selfishness on which that kind of attitude rests. Kant is again attacking a principle of self-love. Not only does the indifferent man need the help of others if the situation is reversed, by virtue of his endorsement of selfishness he is committed to wanting the thing he needs all over again. He could not now renounce inclination and generously decline the help of others even if it were psychologically possible. His way of thinking is essentially ungenerous.[76] This example helps us to understand how general happiness can be the result – but not the *raison d'être* – of the universal adoption of moral maxims.[77]

¶ IV 423.36 The general criterion of moral impermissibility consists in a maxim's generating a 'contradiction in the will', illustrated by the third and fourth examples. In addition, *some* maxims – those that violate strict duties – cannot even be *thought* as a universal law of nature. They generate what is commonly called a 'contradiction in conception': we cannot even conceive of a world in which this maxim universally determines the actions of human beings, 'far less' – Kant now adds – could one '*will* that it *should* become such'. This explains the primacy of perfect duty. Beneficence and other imperfect duties apply only within the limits set by perfect duty: one is not, for example, allowed to make a false promise to obtain the means to help others. A 'contradiction in conception' is the strict criterion that affects only some cases of immoral action, whereas all morally illegitimate maxims create a 'contradiction in the will'. The actual formulation of the categorical imperative accordingly always uses the more general criterion of being able to *will* a maxim as a universal law. A categorical imperative of strict duty reads as follows: act only in accordance with that maxim through which you can at the same time *think* that it become a universal law (of nature).[78]

76 Note that this duty, like the previous one, cannot be reduced to a piece of pragmatic advice. The person in Kant's example is not urged to help others because he thinks that as a matter of fact others might not help him in the future if he does not help them now. He is asked to take up an impartial, moral point of view. The scenario is entirely hypothetical.

77 See the way the 'selfish maxim' is refuted in the *Metaphysics of Morals*, VI 453.5–15. Kant's rejection of the sentiment of indifference ('What's it to me?') is directed at the proverbial lack of compassion displayed by the Stoic sage; see *Metaphysics of Morals*, VI 457.10–12, and the lectures on ethics, *Collins*, XXVII 421.25.

78 There are two reasons to doubt Hartenstein's conjecture *Ableitung* (derivation). *Abt(h)eilung* (classification), as the original editions have it (IV 423.37), should probably be retained as the *lectio difficilior*. First, Kant explains in this and the following paragraph how recognised duties – the prohibition of suicide and lying promises as well as the commands to further one's talents and to help others in need – can be distinguished by the modi operandi of the

¶ IV 424.15 From the 'kind of obligation' (strict or wide) we now turn to duty's 'object' (the agent him or herself or other agents). All duty needs such an object; it must be a duty to someone. In transgressing a duty, Kant says, we find that 'we do not really will that our action should become a universal law, since that is impossible for us' – as has just been shown – 'but that the opposite of our maxim should instead generally remain a law, only we take the liberty of making an *exception* to it for ourselves (or just for this once) to the advantage of our inclination' (IV 424.16–20). In the case of action contrary to duty, the agent excepts himself from a law he otherwise wants to remain intact. The standard from which he excepts himself is thus implicitly acknowledged. In other words: the agent employs double standards, he arrogates to himself a special status, he indulges in an inclination and by this very act denies that others have a right to do the same.[79] *That* is what makes his attitude immoral.

Like the preceding examples, the present paragraph makes it quite clear that Kant's principle concerns the conflict between reason and inclination.[80] Violating a duty to the self one wants to except oneself from a universal law *just for this once*; violating a duty towards others one intends to make an exception from an otherwise generally valid law *just for oneself* – and also just for this once because the timelessness of a maxim must be satisfied in the case of duties to others too. This time Kant does not draw the conclusion that one type of duty (duties to others) overrides the other (duties to the self). However, in the

categorical imperative, i.e. he offers a systematic classification of types of duty appropriate for his present purpose (*beliebig*, IV 421 fn.). Secondly, Kant would have used the old spelling of *Abtheilung*, with the additional h, which makes the mistake less likely than it would be today. Kant does, however, say that individual duties can be 'derived' (*abgeleitet*) from the general formulation and the second variant of the categorical imperative at IV 421.10 and IV 429.8 respectively. Retaining *Abtheilung* does little to moderate the theoretical ambitions of Kantian ethics.

79 In an influential paper, T. Pogge has argued for an 'atypical' reconstruction of Kant's universalisation test. He argues that we must ask ourselves whether we can will that all agents can be *permitted* to adopt the maxim under review, not whether we can will that they *actually* adopt it ('The Categorical Imperative', in *Grundlegung zur Metaphysik der Sitten. Ein kooperativer Kommentar*, ed. O. Höffe (Klostermann, 1989) p. 173). Pogge's reading rightly emphasises the aspect of permissibility, but it is quite compatible with, and even implies, the conventional interpretation Pogge rejects. According to Kant, when inclination urges us to take a certain course of action, the maxim of which is to be put to the test, we face the choice as to whether moral norms should confine our options. If we then decide to flout morality, we must indeed, on pain of contradiction, *allow* all other agents to adopt our inclination-based maxim in relevantly similar situations, i.e. when they are similarly inclined. If so, *they* are urged by *their* inclination to adopt the illegitimate maxim as we are urged by ours, and they *will* adopt it once moral constraints have been removed.

80 In a 'groundwork', there is no room for a discussion of conflicts between different rational grounds; the topic is broached only briefly even in the *Metaphysics of Morals* (VI 224).

Metaphysics of Morals Kant argues for the idea that duties to one's own self are philosophically prior to duties to others because without duties to the self there would be no duties at all (VI 417–18). The most plausible interpretation of this enigmatic thesis is that all duties, duties to others included, are in part duties to the self in the formal sense that they are owed to the binding self, which in an autonomous ethics is to be found in the agent. This fits in well with the idea that all duties rely on temporal universality whereas only duties to others involve universalisation that ranges over all agents.[81]

5 *Interlude*

¶ IV 425.1 The following five paragraphs serve a twofold purpose. First, Kant recapitulates and clarifies what has been established so far. These pages show all the characteristics of a Kantian closing section like the 'concluding remarks' of Section I, and indeed they conclude the part of Section II that roughly coincides with the argument of that section. Secondly, Kant makes the transition to the more metaphysical part of this section, which substantially goes beyond what was said in the previous one. He begins to focus on moral beings as such, and the laws that govern their will. To be laws of metaphysics, they must be laws of *something*, and this something – a pure will – is not yet in sight. This ultimately leads to the discovery of autonomy, the concept of morality required for the justification of the categorical imperative in Section III.

The argument resumes at IV 427.19 with the derivation, from a definition of the will, of a new variant of the categorical imperative. The present paragraph contains a summary of what has been achieved so far. We now possess a clearer account of the meaning of morality, but not proof of its validity (see IV 419.36–420.11). Kant's dictum that a practical law commands 'of itself, absolutely and without any incentives' (IV 425.9–10) must be interpreted quite literally: no incentive is needed for the law to apply to the human will as a command. The will imposes this law upon itself without any external inducement; and an incentive is then created, in recognition of its authority: reverence *for* the law.

81 See N. Potter, 'Duties to Oneself, Motivational Internalism, and Self-Deception', in *Kant's Metaphysics of Morals*, ed. M. Timmons (Oxford University Press, 2002), 371–89, at p. 376, and my 'Kantian Duties to the Self, Explained and Defended', *Philosophy* 81 (2006), 505–30, at pp. 512–15.

¶ IV 425.12 Kant repeats his warning that morality cannot be derived from facts about the physical nature of human beings, lest the law-of-nature formulation be misinterpreted along naturalistic lines (see IV 389.24–35, IV 406–12). Morality is robustly independent of anything natural, particularly inclination.

¶ IV 425.32 The project of grounding a pure moral philosophy is incomplete because we do not yet know what, if anything, it is founded on.[82] God and nature ('heaven' and 'earth', IV 425.33–4) can be ruled out. Moral philosophy must be the 'sovereign empress'[83] (*Selbsthalterin*, IV 425.35) of her own rational laws, not merely the herald of the laws of nature, and thus cannot rely on anything natural or empirical. But we are not yet in a position to appreciate the solution: that in moral judgement we conceive of ourselves as members of a purely intellectual world, which is the ground of our moral sovereignty.

¶ IV 426.7 Kant once again warns us not to mix up the rational and the empirical in moral theory and practice. He alludes to the mythical story of the Thessalian king Ixion, whom Zeus tricked into embracing a cloud rather than his wife Hera. With this cloud, Ixion fathered the first centaur, i.e. a 'bastard patched up'[84] – to use Kant's phrase – of human and equine limbs; see, for example, Pindar, *Pyth. Ode*, 2.21–50.[85] IV 409–10 above suggests that he had the theories of his philosophical rivals in mind. They seem attractive to the untrained eye and may even contain a grain of truth, but those who have beheld 'virtue in her true form' easily recognise them for the absurd concoction they are.

82 The 'standpoint' (*Standpunkt*, IV 425.33) of moral philosophy is said to be unfortunate or uncertain – surely an allusion to the doctrine of the two 'standpoints' developed in Section III (IV 450–3).

83 A literal translation of the Greek αὐτοκράτωρ ('autocrat'), the title attributed to the Russian Tsarina. Cf. *Regiererin* (female ruler) at IV 395.1. Of all the English translators, only Abbott ('dictator') gets it right. In the same spirit, reason is said to 'dictate' (*diktiert*) principles a few lines further down. Denis, presumably to avoid the sinister connotations the word has acquired since the late nineteenth century, substitutes the rather pallid 'director', which obscures Kant's – and Abbott's – point. Ellington's 'author' is problematic because Kant does not think that philosophy or reason devise the moral law. 'Sustainer' is quite wrong.

84 See *Prolegomena* IV 257.34 for the worry that the concept of causality might similarly be the illegitimate offspring of the imagination, rather than a child of reason. For a discussion of Kant's preoccupation with genealogical metaphors within his critical project see I. Proops, 'Kant's Legal Metaphor and the Nature of Deduction', *Journal of the History of Philosophy* 41 (2003), pp. 219–21.

85 See *Mrongovius* II, XXIX 626, and, on the purity of the representation of virtue, R 6917, XIX 206.

Fn. IV 426 Kant emphasises the value of moral action, which is all the more obvious if one ignores its useful consequences. All human beings capable of using their rational faculties can easily come to see this, even a hardened villain (see IV 454.21–2). Is Kant contradicting his strategy, envisaged in Section I, to emphasise the usefulness of morality in order to attract attention to the value of the good will (IV 394.27–31)? He is not. The test proposed in the present footnote is intended for 'experts', as Kant called them in the earlier passage, who make use of their own rational judgement. Usefulness can never be more than a tool or bait that loses its attraction as soon as one progresses on the path to a metaphysics of morals. In moral matters, everyone can be an expert.

¶ IV 426.22 Kant now starts preparing the ground for the second variant of the categorical imperative. Unlike the general formulation and the first variant, the second variant explicitly connects the concept of morality with that of a rational being (namely the human being, a rational nature, which we must consider an end-in-itself by virtue of its moral faculties; see IV 428.21–33). This marks the beginning of a rudimentary metaphysical investigation, as announced in the title of Section II, which leads to the notion of a moral commonwealth and ultimately to an exposition of the concept of autonomy. It is metaphysical because its focus is now on beings, rather than just laws whose origin is unknown; in particular the non-empirical nature of moral beings and its laws.[86] We are also finally leaving behind considerations of prudence, skill and other factors specific to human psychology.[87]

6 *The second variant: rational creatures as ends-in-themselves*

a. Derivation of the 'formula of humanity as the end-in-itself' from the concept of a will

¶ IV 427.19 Like the argument leading to the general formulation and the first variant, the derivation of the second variant formulation starts

86 Kant somewhat paradoxically suggests that there are two parts of the 'doctrine of nature' (*Naturlehre*), one that is natural – in the sense of 'empirical' – and one that is not.

87 Surprisingly, Kant seems to regard the choice of maxims as part of the 'empirical doctrine of the soul' (IV 427.10). If the generation and adoption of maxims could be completely explored empirically it would not be an expression of our freedom of will. Kant presumably wants to say that in the empirical character of a person there are regularities that correspond to, or are expressive of, freely chosen maxims.

with a definition of the will as a capacity of rational beings to act in accordance with laws (cf. IV 412.26–8). However, instead of taking the formal route of distinguishing various kinds of law a human will can follow, Kant now turns to the matter of willing, or its 'end' (cf. V 58.37). The second difference is that, as announced in the previous paragraph, Kant now emphasises the validity of moral laws for *all* rational beings, a metaphysical thesis first introduced in the Preface (IV 389.13–16) and taken up again as the premise of an argument for the freedom of rational beings in Section III (IV 447.26–7).

Unfortunately, the subsequent distinctions are much less clear than those that follow the previous definition. For instance, the thought that 'ends' should be objects in the world and therefore 'objective', rather than the subjective purposes an agent intends to realise, is rather peculiar. It is in this literal sense that Kant speaks of the end as the *objective* ground[88] of self-determination (IV 427.22). Of course, such an objective ground is also the foundation of an agent's *self-determination* in the sense that all merely subjective grounds can at best effect external determination of the will by inclinations.

There are three considerations that help us to explain Kant's usage. First, the German word for 'end', *Zweck*, originally signifies the bullseye of a target that an archer intends to hit, i.e. something external to and independent of the agent and 'objective' (object-like) in this sense. An end is, secondly, that for the sake of which an action is performed (to continue the simile of the marksman, it is that at which an action is aimed). From the first-person point of view of the agent, ends can easily be processes or external objects – money or satisfaction can be the end of one's work and in this sense be the 'objective ground of self-determination'. In the case of a benevolent moral act we frequently say, in a Kantian spirit, that the agent did it for the sake of the other person. Thirdly, the idea that ends are some kind of thing or object is also suggested by the opposition of means and ends.[89]

88 H. J. Paton, in his translation, ventures to substitute 'subjective ground' for 'objective ground'. The idea is that all ends must be subjectively chosen and can never be externally imposed. On this picture, some ends – those that are given solely by reason – become objective grounds as well. The second variant then marks a change of emphasis from what is objective in moral volition (the law) to that which is subjective (the end). The argument leading up to the third variant at IV 431.9–18 below admittedly makes Paton's emendation more attractive.

89 Note that there is little temptation to 'subjectivise' or internalise means – presumably because as agents we value means as purely instrumental and avail ourselves of them as we need them, whereas ends are properly adopted. We naturally call ends 'our own', but we do not – in the philosophically relevant sense – speak of 'our means'.

The argument of the present paragraph rests on a difficult terminological distinction between two perspectives on human action (IV 427.26–7): the subjective ground of our desiring something is called its 'incentive' (*Triebfeder*); by contrast, the objective ground of why something is willed is its 'motivating ground' (*Bewegungsgrund*). Kant seems to be assuming that although all human action involves both incentives and motivating grounds, some actions – those in pursuit of merely subjective ends – *rest* on incentives, whereas with others – which involve objective ends – it is motivating grounds that take the lead.[90] *Subjective ends* owe their status as ends to a subjective pro-attitude on the part of the agent, whereas *objective ends* are valid for all rational beings.[91] In the latter case, motivating grounds influence the will more by way of attraction, effected by objective, universal ends when recognised and judged by reason. This distinction between two radically different kinds of 'end' is taken up again at IV 437.23–30, where Kant distinguishes ends that are 'to be effected' from 'independently existent' ends.

Practical principles (imperatives) are defined as 'formal' if and only if they do not depend on incentives, i.e. sensuous, natural desires or motives (IV 427.30–2). These imperatives can therefore take 'grounds of motivation' and objective ends into account, their formality notwithstanding. A 'motivating ground' is not itself a motive in the modern psychological sense of the word – some kind of desire – but rather an objective ground that triggers a novel interest or motive in the agent. Kant intends to distinguish between principles that depend on certain subjective conditions and those that do not. In the same spirit, Kant calls *reverence* the 'incentive' of moral action in the *Critique of Practical Reason* and at IV 440.5–7 below – even though it is not, of course, a conventional incentive. It does not influence the command of the will but is first generated by it. As the last sentence of the present paragraph indicates, we are crossing – if on a different route – the

90 There is thus no reason to think that Kant had to 'broaden' his conception of incentives in his later writings to include motivation triggered by pure reason (see Wood, *Kant's Ethical Thought*, pp. 360–1). Kant is saying that subjective ends *rest* on incentives, whereas objective ends do not. In the *Groundwork*, as in the *Critique of Practical Reason*, reverence – a purely rational source of motivation – is an 'incentive' – in fact the only incentive that can make actions morally valuable; see IV 440.5–7 and V 81.20–5.

91 To quote the sentence in full: 'The subjective ground of desire is an *incentive*; the objective ground of willing is a *motivating ground*; hence the distinction between subjective ends, which rest on incentives, and objective ends, which depend on motivating grounds that hold for every rational being' (IV 427.26–30). Like Wood, Zweig and Paton, contra Mary Gregor, I consider the last relative clause to be defining and the two previous, parallel relative clauses to be non-defining, as the translation indicates. The original German is ambiguous.

familiar territory of hypothetical imperatives, which rest on subjective ends, and the categorical imperative, which calls for an objective end.

¶ IV 428.3 A moral action according to formal principles must be performed for the sake of some end, or from reverence for such an end, even if it cannot rest on merely subjective ends. We now enter the search for an objective end. This is an end that by itself does not merely possess subjective value for an agent, but absolute value that every rational agent is bound to acknowledge, whatever (subjective) ends he or she may happen to have.

One thing, however, is already clear. An objective end cannot be the kind of end that is furthered or brought about by an action. Such a 'teleological end' is entirely unsuitable because, as demonstrated in Section I, the external effects of our actions are never entirely within our control. The end of morality must be special. It must be different in kind. Kant is clearly conscious of the philosophical and terminological extravagance of his thesis. But as the appreciation of moral ends in this sense does not depend on subjective preferences, Kant is not abandoning his project of a purely formal ethical theory, however paradoxical it may at first seem.[92]

The search for an absolute end can be illustrated as follows. Something that we consider an *end* from one perspective can easily become a *means* to a further end. If the object of my inclination is a cup of coffee, and my inclination makes it seem desirable to me, other objects like coffee beans, grinders, kettles, pots, filters and water are *means* that I employ to realise this end. The cup of coffee, as the end or object I wish to bring about, contrasts with the subjective incentive within me that prompts my value judgement. I have achieved my end when there is a cup of steaming hot coffee on my desk in front of me. Importantly, that which is the end of my putting the kettle on, grinding the beans, etc., the cup of coffee, is itself a means towards a further end: my drinking it – and there is nothing wrong with now treating a cup of coffee as a *mere* means, or for that matter a coffee grinder, a kettle, a bag of coffee beans. Once I have brought about a certain end it can become a means towards a further end. If so, not much of its initial status as the end of my action remains. We can play the same game over and over

92 See also the corresponding distinction in Kant's summary of the second variant at 437.21–30 below, where its formalism is reasserted.

again. The end of my putting coffee beans in the grinder is the product, the ground coffee, which in turn becomes a means once the water has come to the boil. Some of the objects called ends then are merely subjectively good; amongst them all the ends that can be procured or brought about by my agency. Others are objectively good in the sense that they must be considered valuable impartially and universally. Our capacity to act for the sake of something that is not backed by inclination depends on our acknowledging the existence of ends of the latter kind. What kind of being could have this elevated status? Kant is now approaching the question of what kind of object *never* ought to be used merely as a means but must rather be treated in a different way, as an end-in-itself.

¶ IV 428.7 In the course of this paragraph, Kant argues for his favourite candidate for the status of an objective end – 'man [*der Mensch*] and in general every rational being' (IV 428.7–8) – by systematically excluding all other possible contenders. Kant prefaces his survey by saying that such an end must always be treated as an end-in-itself and never be regarded merely as a means towards other, merely subjective ends.

Three candidates are quickly rejected. First the objects of inclinations, which possess value for us only if we are favourably disposed towards them; secondly, inclination itself as the condition of the subjective value of these objects. The idea behind this regressive move is simple. Some object such as a cup of coffee has value for me only given that I like and desire coffee; but if the object of my desire is not absolutely valuable one might consider the condition that *makes* it so: inclination. But Kant also rejects this second candidate. Inclinations are the reason why we require things (they are the 'sources of need', IV 428.14–15). If possible, a rational creature would wish to be without them.[93] One method of ridding oneself of an inclination – in fact the common one – is to see to its satisfaction. Arguably, the very fact that we possess inclinations that require satisfaction is a kind of deficiency because it makes us less self-sufficient. Yet would we really choose to be without inclinations if we were given the choice? We would certainly opt out if the alternative was leaving them unfulfilled. As to the stronger claim

93 This argument should not be seen as an expression of Kant's hostility to the sensual side of human nature. He is not calling for the eradication of inclination, which must merely be cultivated and subordinated to reason; see e.g. *Religion*, VI 58.1–6. Being completely free from inclination can never be more than a wish.

that we wish to be rid of inclination altogether, Kant would presumably argue that any desire to keep inclinations – perhaps for 'the good things in life' – must itself be based on inclination. If we could simply wish them away, we would not miss them or judge our life to be any worse.[94]

Thirdly, Kant considers non-rational animals. He thus moves from entities that can be brought about by the human will, their value and its conditions, to independently existing objects. Animals cannot be ends-in-themselves or possess absolute value, which is borne out by the fact that, unlike persons, they do not evoke reverence (see *Critique of Practical Reason*, V 76.24–7). Kant calls them 'things' and declares that human beings have the right to use them for their own purposes, as mere means. As the second variant indicates, duties must always be directed at persons, either oneself or others. Consequently, there can be no direct duties towards animals. At this point, Kant does not provide arguments for his contentious classification of animals as things (but see, for example, *Metaphysics of Morals*, VI 443).

Once non-rational animals have been excluded, only the fourth candidate remains: persons – rational independently existing beings that unlike things cannot be exchanged arbitrarily. Only persons are 'ends-in-themselves'.

¶ IV 428.34 Kant now draws his conclusions from what has been said so far. The peculiar end that is required for the categorical imperative to be possible is man, a human being (or any rational creature like it). The ground of the categorical imperative is: 'Rational nature exists as an end-in-itself' (IV 429.2–3).

This statement is easily misunderstood. 'Rational nature' appears to refer to a quality or capacity that human beings possess, but on closer inspection this reading is untenable. When Kant says that *die vernünftige Natur* exists as an end-in-itself, *Natur* is synonymous with 'being' of a certain kind (or nature): a rational creature, a person. Moreover, the

94 See the 1784–5 lectures on ethics: 'The thing towards which we have an inclination pleases us, but the inclination in itself does not please us, for otherwise [without the inclination in question] we should not have so many needs' (*Moral Mrongovius* II, XXIX 610.6–8); see also *Critique of Practical Reason*, V 117–18, where Kant declares the contentment that accompanies moral conduct to be due to a concomitant independence of blind inclination. In the *Critique of Judgement*, the elevated status of the satisfactions of beauty is justified on the grounds that it is wholly disinterested. This kind of aesthetic approval neither presupposes nor produces a need (to act); thus it is the only kind of judgement that is completely free; see V 209.21–5. On needs see also IV 413–14 fn. above.

definite article is used to express a general statement: *any* rational creature *as such* is an end-in-itself. The first point is less than obvious to today's readers of Kant's original German;[95] both are almost completely obscured by the veil of translation.[96]

Having argued for his favourite candidate, Kant offers a second variant of the categorical imperative. Owing to our special status, we should always use 'humanity' in ourselves as well as in others at the same time as an end and never merely as a means (IV 429.10–12; see second *Critique*, V 87.13–30 and V 131.20–4). In this formulation of the categorical imperative, 'humanity' refers to the rational capacities of human beings. We should respect and further, 'within' ourselves and others, the element or activity which makes us human.[97] In the lectures on moral philosophy, Kant expressly distinguishes a 'villain' (*Bösewicht*; cf. IV 454.21–2) and his 'humanity' (*Menschheit*) – we can be pleased with the latter even if we are rather less than pleased with the former, the person as a whole (*Collins*, XXVII 418.17–19).

Three further notes on the new formulation. First, Kant is not saying that we should never use other human beings as means – we inevitably do. In fact, our treating others as means is an important source of duties towards them. When I eat out, I use the chef and the restaurant's waiting staff as means towards my own purpose of having a nice meal. Kant is saying that one should not use them – or anyone else – *merely* as a means, i.e. regard them as mere tools or instruments. Whether we pay

95 However, *Natur* in this sense is still occasionally used in contemporary German, in phrases like *eine heitere Natur* (or *Frohnatur*), a naturally cheerful person.

96 A review of passages from Section II confirms this reading. First, right at the beginning of the section Kant queries the role of reverence, which seems to be bound up with the human condition, as a 'universal precept regarding every rational nature' (IV 408.21–2); note the quantifier. Kant uses the phrase 'rational being' throughout that paragraph. *Vernünftige Natur* is little more than a stylistic variation. Secondly, after illustrating the second variant with the four usual examples, Kant asserts that '[t]his principle of humanity and in general of every rational nature as an end in itself . . . is not borrowed from experience' (IV 430.28–431.2). Here, 'rational nature' sounds so odd that Paton decided to render it 'rational being'. He could, thirdly, have done the same a few pages further down: 'Autonomy is the ground of the dignity of a human and every rational nature' (IV 436.6–7). Fourthly, 'a rational nature is distinguished from all other natures [*den übrigen, sc. Naturen*] in that it sets itself ends' (IV 437.21–2). In this last quotation, as elsewhere, the definite article (*die vernünftige Natur*) is used with the singular to indicate the generality of the statement. *Natur* is not a specific quality, but rather a being or object of a certain kind. This usage, or indeed the universal quantifier, would hardly make sense if Kant was not talking about an individual creature but about 'rational nature' in the abstract, as a capacity. Both the German and the English words have their roots in the Latin *natura*, as used in similar philosophical contexts.

97 Anticipating the later distinction between animality, humanity and personality in *Religion within the Limits of Reason Alone* (VI 26–8), Kant is perhaps referring to our rational faculties in general, not just our moral capacities. The range of duties derived from the categorical imperative seems to confirm this.

due attention to the status of human beings as ends-in-themselves has practical consequences. Our attitude, be it careless or respectful, will affect our behaviour. It is a mark of moral action towards others that they could rationally assent to one's principles if they were known to them.

Secondly, the injunction to treat human beings 'always as an end, never merely as a means' does not consist of two independent criteria that must always be jointly applied. They are two distinct criteria, but they overlap in cases of strict or 'necessary' duty only (cf. IV 430.10–13). In that case, not according someone his proper value as an independent, objective end is a direct consequence of treating him as a mere means; which is not the case with regard to wide or 'contingent' duty.[98] If I refuse to help those in need, I fail to pay due attention to their status as fellow human beings and thus act contrary to meritorious duty, but I do not use them as instruments towards my own ends. Like the first variant, the second contains both a narrow and a broad criterion of moral action. Both apply in the case of narrow duty, whereas only the weaker applies in cases of wide duty. Again, this must be the reason why – as Kant now puts it – 'necessary' duty takes precedence over 'contingent' duty. 'Act in such a way that you treat humanity in your own person, as in the person of any other, never merely as a means' would be the categorical imperative of 'necessary' duty.

Thirdly, equating the second variant and the first it is not entirely clear how Kant conceives of the mutual relation of ends and laws, matter and form. Does the recognition that we as well as others are ends-in-themselves force us to act in accordance with strictly universal formal laws? Or does the moral law within us first enable us to recognise other persons as ends-in-themselves, which then in turn makes us act in accordance with universal standards? The second approach has the advantage of invalidating the objection that the second variant

98 We should note that the transition to the second variant of the categorical imperative is marked by a terminological change. Kant no longer speaks of 'perfect' (or 'strict') and 'imperfect' (or 'wide'), but rather of 'necessary' (or 'owed') and 'contingent' (or 'meritorious') duty. Wide duties are 'contingent' in that to enter into force as token duties they depend on a particular occasion, such as in the case of beneficence the predicament of another human being. They are not, of course, contingent on the agent's inclination towards dutiful action. That would turn these duties into hypothetical commands. Also, that they are called 'meritorious' should not be taken to mean that they are in any sense less obligatory once they apply. (Any token duty makes the action required necessary; see the official definition of duty at IV 400.18–19.) However, the person who receives our help does not have a right to our assistance, nor do we wrong him if we refuse to help. He does have rights, which we owe him, such as the rights not to be deceived, deprived of his property or murdered. He is obliged to be grateful for our help.

implicitly relies on moral norms, and also shows its limitations; form is still primary.

As in the previous case, we should ask why Kant offers us this second variant. He may be thinking of it simply as another step towards the metaphysical ideal of self-legislation in a moral commonwealth, the central notion of the third variation, which is declared to arise from a combination of the previous two (IV 436.8–26, cf. IV 431.9–13). Does Kant have a specific target in mind? He was certainly conscious of the philosophical novelty of the notion of a person as an 'objective' end-in-itself that is radically different in kind from what philosophers normally regard as an end. At IV 436.12–13 we learn that all three variants of the categorical imperative rest on 'a certain analogy'. In the present case, it is that of treating humanity as a very special kind of 'end'. Kant is likely to be aiming at an ethical theory of conventional ends, i.e. classical teleology.

Over and above that, the motivation of his transition to the formula of humanity as an end-in-itself is unclear. He may simply be thinking of a teleological world order in which man is the final end for the sake of which everything else exists. There is, however, one particularly striking parallel in the history of philosophy: the characterisation of the final and complete end as that for the sake of which every thing is done and which in turn is done for the sake of nothing else. This definition stayed in place throughout antiquity, and the answer was always the same – εὐδαιμονία – while the substance of ancient happiness underwent major changes.

Compare the following passage from the *Nicomachean Ethics*:

> Now we say that what is worth pursuing for itself is more complete than what is worth pursuing because of something else, and what is never desirable because of something else is more complete than those things that are desirable both for themselves and because of it; while what is complete *without qualification* is what is always desirable in itself and never because of something else. Happiness seems most of all to be like this; for this we do always choose because of itself and never because of something else (1097a31–b2).

The similarity is striking. If Kant has Aristotelian teleology in mind (there is little indication that he does),[99] the second variant would be

99 The only element of Aristotelian ethics with which Kant is (superficially) acquainted is the doctrine of the mean; see e.g. *Metaphysics of Morals*, VI 433 fn., and the *Collins* lecture notes, XXVII 277.5–6. Kant possessed a complete edition of Aristotle's works in Greek and Latin, but the *Nicomachean Ethics* is not on his list of 'Aristotle's best books': 'Logic, Rhetoric, Natural History' (R 1635, XVI 57, early 1750s).

suggesting that classical teleology was right about the structure of the final end, and that ethics can indeed be done in terms of ends, but that believing it to be happiness unfortunately the Ancients misidentified the substantive final end. The final end is, in fact, humanity.[100]

Fn. IV 429 Kant refers to the two sentences preceding the asterisk, i.e. the idea that *every* human being regards him or herself as an end-in-itself. He uses this argument to support his metaphysical thesis that 'rational nature exists as an end-in-itself', the ground of the objective principle of the second variant. The fact that one regards both oneself and other rational beings in this fashion leads to duties towards both oneself and others, which is why these two objects of duties are then distinguished in the text above. The subjective and objective necessity of this self-ascription is grounded in the membership of every human being in an intellectual world, introduced only at IV 450.30. The burden of proof of the new variant also rests on the justificatory story told in Section III.

b. Application of this formula to the four examples of duty

¶ IV 429.14 Kant re-uses the four examples of widely recognised moral duties first introduced at IV 421–3.

¶ IV 429.15 The first illustration. Morally, no-one is at another person's arbitrary disposal, not even at his or her own. Kant argues that someone who kills himself because he is sick of life treats himself as a mere means towards escaping his burdensome state. Such actions are said to contradict the idea that human beings are ends-in-themselves. Kant's prohibition extends beyond 'total' suicide to self-mutilation and self-corruption: we must never use a human being, ourselves included, as a mere instrument towards our inclination-based purposes. Limbs and organs are parts of human beings.

However, Kant makes clear that the categorical imperative does not prohibit all kinds of intrusive physical operation (see the parenthetical remark at IV 429.25–8). As always, the attitude of the agent is the decisive factor. The target of the first example is again a 'principle of

100 In fact, even G. A. Tittel acknowledges the principle that one should always treat humanity as an end; but then, anticipating R. M. Hare's and D. Cummiskey's work, he asks whether this does not amount to the very principle that Kant rejects: 'the principle of happiness, i.e. of self-love and beneficence' (*Über Herrn Kant's Moralreform*, p. 46).

self-love', i.e. a maxim proposed by inclination (cf. IV 422.7). In the *Metaphysics of Morals*, Kant says that even having one's hair cut is 'not entirely without guilt' if it is done for the sake of selling it to a wig maker, rather than for reasons of personal hygiene (see VI 423.13–16), which is followed by a brief discussion of apparently legitimate cases of suicide. None of these complications form part of a 'groundwork' of the metaphysics of morals.

¶ IV 429.29 The second illustration. Kant argues that someone who intends to obtain a loan by means of a false promise violates the status of the lender as an end-in-itself, in particular: he treats the lender solely as a source of funding, as a mere means or instrument, without due attention to his or her legitimate interests. A fortiori, the person lied to cannot share the end of the liar – of being swindled out of his or her money. The two cannot co-operate as equal partners or pursue common ends. A necessary duty such as that not to lie to others is owed (IV 429.29), and the other person has a corresponding right not to be thus treated (IV 430.5).

Fn. IV 430 Kant attaches his footnote on the difference between categorical imperative and Golden Rule to the illustration that is closest in spirit to the latter. They are obviously similar in that they are both formal in the sense that they do not depend on any particular purpose or intention; and they both operate on a notion of practical assent. The Golden Rule is stated negatively: *quod tibi non vis fieri, alteri non feceris* (see Thomasius, *Instit. jurisprud. divinae*, I, 4, § 18; II, 3, § 21), i.e. do not treat others as you would not be treated (see Tob. 4, 5). The fact that one does not wish to be treated by others as a mere means towards their ends – and thus have one's rights violated – gives any rational being a reason to treat them in the way he or she expects to be treated: decently.[101] Moreover, Kant is obviously right when he says that duties to the self and duties of beneficence towards others, cannot be derived from the Golden Rule as stated; whereas the argument that a criminal could use the Golden Rule against his judge, arguing that after all the judge would not wish to be imprisoned by the criminal, seems artificial. Can the judge not grant that he would have to be treated in like manner if he was guilty of the same crime, which he is not? Even if, if

101 Kant endorses the same formulation of the Golden Rule in the *Mrongovius* II lectures on moral philosophy as a second-tier principle in a 'kingdom of ends'; see XXIX 610.36–7.

he was guilty, he would not like being punished? [102] Nor does it occur to Kant that his new formulation of the categorical imperative might be vulnerable to similar uncertainties and objections.

¶ IV 430.10 The third illustration. To establish a duty to be industrious Kant once again introduces teleological principles, and again the question arises whether he is justified in doing so. Our capacities or predispositions (*Anlagen*) are said to be 'part of the purpose of nature with regard to humanity in our subject'; and letting our talents rust is supposed to be compatible with sustaining humanity but not with advancing its cause. No further argument is given. Presumably, the development of one's talents seems to acquire value because, as rational creatures, we are capable of testing and pursuing ends independently of our inclinations.[103]

¶ IV 430.18 The fourth illustration. The happiness of others is an object of duty. All human beings naturally wish to be happy. Again, it cannot be a desire based on inclination alone that calls for the assistance of others, but rather the fact that human beings – rational creatures, objective ends that are the subjects of other ends – are capable of performing actions that are objectively (morally or prudentially) good. That is why taking someone seriously as an end-in-itself implies taking on his ends as one's own, to co-operate with him, to help him. This duty, like all others, rests on the *capacity* of practical reason as the ground of human dignity – ignoring the fact that we often behave imprudently or immorally. It is not entirely clear how far duty, especially positive duty,

102 See R 7994, XIX 576: 'What is the foundation of the *potestas puniendi*? Not the assent of everyone to be punished, but the will of everyone to punish everyone else.'

103 The first sentence of this paragraph contains a veiled description of the two kinds of moral flaw involved in acting contrary to the second variant. These flaws correspond to the contradictions 'in conception' and 'in the will' of the first (see IV 424.1–14). Kant says that in the case of meritorious or wide duty to the self it is *not sufficient* if our action does not contradict (*widerstreite*, IV 430.12) humanity in our own person as an end-in-itself, i.e. that we do not use ourselves merely as a means. Our action must *also* be in harmony with the special status of humanity (*dazu zusammenstimmen*, IV 430.12–13). There are therefore two ways of failing to treat people appropriately: first, to use them as instruments, in which case a fortiori they are not treated as ends-in-themselves; secondly, not positively to further their ends. Consequently, at IV 429.17–19 (first illustration) and IV 429.31–3 (second illustration) both criteria are used, whereas at IV 430.10–17 (third illustration) and IV 430.23 (fourth illustration) only the general criterion of a positive coincidence with humanity as an end-in-itself is applied. This makes good sense. With regard to duties to others, two people cannot be equal collaborators when one of them treats the other as a mere means to his own purposes; but *not* treating another like an instrument falls short of mutual collaboration, of sharing the other's ends. Negative duty is primary, but we are morally bound to go beyond mere non-interference.

extends to morally bad people, but there is certainly no obligation to support their vices; see *Metaphysics of Morals*, VI 480–1.

¶ IV 430.28 To conclude the discussion of the present formulation Kant again, if rather more briefly, tries to distinguish his approach from popular ethical theories (cf. IV 425.12–31). He now targets eudaimonistic systems, which define the final end in terms of the agent's subjective ends. Kant's 'objective' end of humanity is different in kind from all such ordinary ends. His new formula is not derived from experience; first because of its universality, secondly because the ends in question do not rest on subjective conditions and therefore cannot be empirically investigated. Objective ends rationally *restrict* the choice of all subjective ends. Note that the categorical imperative limits our freedom of *choice* to actions that cohere with humanity as an end-in-itself. Our freedom of *will* consists precisely in our being able to obey the categorical imperative.

7 *The third variant: autonomy in a kingdom of ends*

a. Derivation of the formula of autonomy from the other two

¶ IV 431.9[104] Kant now proceeds to the formulation of the third variant. Unlike the general formulation and the second variant, the 'third practical principle'[105] is not directly derived from a definition of the will. Rather, it combines features of the statements of the categorical imperative presented so far. This much is clear. The details of the argument leading up to the notion of autonomy are rather less obvious. In the light of the summary of the formulations at IV 436.8–28, the following reconstruction would seem to be the most promising.

Any act of volition involves a formal and a material element. The former is the law that – objectively – governs the action, the latter the end that the agent – the subject – freely adopts.[106] In the case of 'practical' – i.e. moral – legislation, the 'rule' (maxim) is such that it can be

104 Like T. Valentiner, I should like to insert a paragraph break at this point.

105 Strictly speaking, Kant should not call the variant formulations 'principles'. All formulations are representations or variants of the same principle; see IV 436.8.

106 When Kant says that the ground of practical legislation subjectively lies in the end (IV 431.12), does he mean (i) that all ends need to be subjectively endorsed by a rational agent? Or (ii) that according to the second variant moral action is directed at ends-in-themselves? Or (iii) that its objectivity notwithstanding the moral law is grounded in an 'end', i.e. the rational being as its subject? This detail of the argument remains obscure. The derivation would be strengthened by Paton's substitution of 'subjective' for 'objective' at IV 427.22. The first formulations would then emphasise the objective aspects of a universal moral

willed as a universal law, as the initial formulations reveal. Moreover, as we learn from the second variant, there are ends-in-themselves which as subjects contain all other ends, and every rational being is such an end. Combining the objective and subjective aspects of the preceding formulations, Kant thus arrives at the 'idea' – a necessary concept of reason that does not permit of an adequate object in experience (A 327/B 383) – that *every rational being* (see the second variant) may be engaged in the promulgation of *universal laws* (envisaged in the general formulation and the law-of-nature variant). Kant's repeated emphasis that *every* rational being is capable of making universal laws points us towards the ideal of a 'kingdom of ends'; see IV 433.16 below.

b. A universally legislative will is independent of all interest

¶ IV 431.19 Like the general formulation and the formula of the law of nature, the third variant – only hinted at so far – serves to dismiss morally impermissible maxims. (The second variant does so implicitly by rejecting maxims that do not pay due respect to the status of human beings as ends-in-themselves.) But even though the formula of autonomy resembles the general formulation, and at IV 434.10–14 even appears directly to follow from it, the explicit introduction of the subject of legislation brings to light the characteristic feature of Kantian autonomy: that by virtue of making universal law every rational agent subjects him or herself to it. The law would not otherwise be universal. The moral law is grounded in the agent's rational self and is completely independent of any external influence. We decide to side with morality *because its law is essentially our very own law*. That it is one's own law from which one exempts oneself in immoral action was at best implicit in the first formulation. Autonomy is further explained in the following paragraphs. A formal discussion concludes Section II at IV 440.14 below.

¶ IV 431.25 Disclosing the source of the moral law the formula of autonomy also clearly reveals what was merely implicit in the preceding formulations of the categorical – i.e. unconditional – imperative: that morality must be independent of all human interests (which rest on

law, whereas the second variant would present morality from the subjective perspective of individual agents. Similarly, the objective determination of volition is associated with the law, the subjective determination with the incentive of reverence (which enables us freely to adopt moral ends into our maxims), at IV 400.32–3.

inclination). The possibility of this independence is only explained in Section III: we are members of a superior intellectual world with its very own laws; see IV 451.24–36.

¶ IV 432.5 The law cannot depend on any interest because the will itself must be the supreme lawgiver. (Otherwise the law would be grounded in some interest, not in the will, as in action from inclination.) However, the will can subsequently *generate* an interest in moral action: reverence for the law.[107]

¶ IV 432.12 In a twofold argument, Kant concludes that by virtue of its unconditional nature the newly discovered principle provides for the human will a categorical imperative; and, conversely, alluding to the derivation of the categorical imperative from its concept (IV 420.24–421.8), that a categorical imperative is a command of autonomy that does not presuppose any interest.

Fn. IV 432 Kant encourages his readers to illustrate the workings of the new variant with reference to the four examples that follow the universal-law-of-nature formulation (IV 421–3). It has already been stated that the second variant leads to the same results; see IV 429.14. The three variant formulations are obviously meant to be practically – but not philosophically – equivalent. The location of the asterisk indicates that Kant thought of the 'principle' to which it is attached as the canonical formulation of the third variant, even though it is not stated as an imperative.[108]

¶ IV 432.25 Kant returns to his critiques of earlier systems of moral philosophy. For all their differences, they make the same fundamental mistake. They ignore the autonomy of the human will. If the law is

107 It is impossible to explain how this happens; see again Section III, IV 458.36–460.7.

108 The expression 'the categorical imperative and its formula' is intriguing. It might seem to refer to the general formulation and the first variant respectively. The four examples illustrate the general formulation only indirectly – they are attached to the law-of-nature formulation; and the general formulation is frequently distinguished as 'the categorical imperative'. The word 'formula' also – de facto – refers to the three re-stated versions of the categorical imperative at IV 436.9 below. Yet this reading is not entirely convincing. The basic statement of the categorical imperative is itself a formula (see IV 413.10); and it would be unfortunate if Kant were speaking of a formula 'and its formula'. In the light of the argument of the main paragraph to which this footnote belongs it seems rather more likely that, as suggested at IV 420.19, 'the categorical imperative' refers to an unconditional command as such, and 'its formula' to one of its earlier formulations (which happens to be the first variant).

not grounded in the will itself, it is conditional and depends on external checks and balances, such as divine sanctions. Consequently, all imperatives will turn out to be conditional, i.e. hypothetical and not really moral at all. Autonomy and heteronomy are first and foremost properties of different types of will, and by extension also of ethical theories that explicitly or implicitly presuppose respective models of the human will. Kant resumes his critique of other philosophical systems at IV 441.25.

c. Self-legislation, morality and the kingdom of ends

¶ IV 433.12 This short paragraph is little more than a promise that the notion of autonomy may lead to another, closely connected one: that of a 'kingdom of ends'.[109] It is made good, after some preparatory remarks, two paragraphs further down.

¶ IV 433.17 Kant defines a 'kingdom' (*Reich*) as 'the systematic conjunction of different rational beings through common laws'. For the purpose of creating such a kingdom, the private purposes of its members – their inclinations and their happiness – are immaterial. (They may realise them within the limits of the morally permissible, but that is not a matter of common legislation.) However, as the kingdom of ends contains both rational beings as ends-in-themselves and the (moral) ends they freely choose, it also contains a formal endorsement of the permissible private ends of others. Every member is called upon to be beneficent and help others in their pursuits. As the second variant indicates, the negative compatibility of the ends of different members is not sufficient (see IV 430.22).

¶ IV 433.26 The ideal of a kingdom of ends now also involves the other persons affected by the actions of agents who apply the categorical imperative. Ends treat their fellow ends – as well as themselves – as they deserve. Kant apparently regards the formulations of the categorical imperative referring to autonomy and the kingdom of ends as statements of one and the same 'third' formula, which may well seem

109 On the respective merits of 'kingdom' and 'realm' as translations of *Reich*, see Paton, *The Categorical Imperative*, pp. 187–8. I agree with Paton that 'realm' is too bland. It does not convey the sense of structure and order that Kant clearly associates with *Reich*, and it obscures the implicit reference to a kingdom of God. On the other hand, 'kingdom of nature' (*Reich der Natur*, IV 436 fn.) may have unwelcome connotations of its own. There is no perfect translation.

surprising. His attitude does, however, make good sense if we give due weight to the fact that autonomy, as first introduced at IV 431.16–18 above, implicitly refers to multiple 'ends' as agents. If we emphasise the aspect of universal legislation, the idea of a systematic, harmonious commonwealth readily comes to mind.

Conflict is excluded from such a realm because the laws on which these agent act are, by definition, universal laws. Kant calls the kingdom of ends an 'ideal' because it cannot be fully realised by finite beings like us. But it is part of the metaphysical story that motivates and inspires us to live up to our autonomy. The reason is that it reminds us of our membership of an intellectual world, officially introduced and defended only in Section III. Our being part of the other world enables us to shape this world according to moral norms. The ideal of a kingdom of ends is realised in this world if everyone follows the laws of the other world in his actions.

Like the previous two variants, the formula of autonomy in a kingdom of ends also serves a rhetorical function. In the 1784–5 lectures on moral philosophy Kant indicates that he considers 'kingdom of grace' Leibniz's expression for what he calls, and in fact is, a 'kingdom of ends' (*Mrongovius* II, XXIX 610.35–6).[110] Of course, Kant does not credit Leibniz with being the first to formulate the categorical imperative, or to believe that the new metaphysics of morals can be rendered dispensable by an older one. He is trying to win Leibnizians over by suggesting that their metaphysics, like Stoicism and teleological ethics, contains a grain of truth worth preserving.

¶ IV 433.34 The kingdom of ends, like the kingdom of grace, is headed by God.[111] Like the sovereign of an eighteenth-century state, he is not subject (*unterworfen*) to the law he legislates, if for slightly different reasons. The head of the kingdom of ends is not exempt from the law. In opposition to the voluntarism of Crusius, he is part of the general moral community of rational beings. But the law applies to him descriptively, not as an imperative. He does not need to bind himself and is not therefore bound (see second *Critique*, V 32.17). Furthermore, human beings are asked to imagine that they subject themselves as well as other

110 See G. W. Leibniz, *Metaphysical Treatise* § 36, *Monadology* § 87 and the *Theodicy* passages mentioned there. Kant's ethics lectures end on the same note. He expresses his hope that there may come a day when the vocation of humanity is fulfilled, a 'kingdom of God on earth' (*Collins*, XXVII 471.33); see also *Critique of Pure Reason* A 812/B 840.

111 We can conceive of the kingdom of ends as the realisation of the kingdom of God on earth, and at least part of the highest good; see second *Critique*, V 128.1 and V 137.2.

agents like them – not the head – to the universal laws of their own making. That is what Kant means when he says that the head 'is not subject to the will of another' (IV 433.37) – which, literally interpreted, is true of *all* rational beings by virtue of their autonomy. The same moral law applies to God, but he is not included in the thought experiment of universalisation.

The idea of a moral sovereign calls for further clarification. According to the *Metaphysics of Morals* (VI 227.10–20),[112] we must distinguish two aspects of legislation. There is (i) the lawgiver or legislator, the one who imposes the law. He is said to be the 'author of the obligation' (*Urheber der Verbindlichkeit*), i.e. he is responsible for the fact that the law applies, and an obligation exists. We have already seen that any autonomous agent must consider himself a lawgiver of his own (moral) laws, and hence the author of the corresponding obligation. But a lawgiver is not always (ii) the 'author of the law' (*Urheber des Gesetzes*), the person who devised the law, i.e. determined its content. There is no author of the moral law because it is not arbitrary. It could not be otherwise than it is. Like the laws of geometry, it has its origin neither in our will nor, for that matter, in the will of God.[113] In addition, a philosophy of religion grounded in reason, not revelation, teaches us that moral laws are endorsed by God as (iii) the 'moral ruler of the world' (*moralischer Weltherrscher*). This is a separate point, prominent in his *Religion* (see, for example, VI 99) but not yet implied by what Kant says about the head of the kingdom of ends in the *Groundwork*. Such a ruler authors neither obligation nor its law, in the sense discussed above. Kant argues that the moral ruler cannot impose laws because that would make all obligation juridical and leave no room for morality.[114] An external authority can impose a command to act in *conformity* with duty; but it cannot command that we do so from a moral *attitude*. That we have to do ourselves. Autonomy is both a sufficient and a necessary condition of morality.

¶ IV 434.1 Members and head alike are autonomous legislators, but whereas the kingdom of ends must be realised by acts of freedom of the

112 P. Kain has recently traced the distinction to R 6513 (XIX 48), a reaction to § 100 of Baumgarten's *Initia*, which he dates 1762–3; see 'Self-Legislation in Kant's Moral Philosophy', *Archiv für Geschichte der Philosophie* 86 (2004), 257–306. See also the account in *Moral Mrongovius* II, XXIX 633.26–634.2, and R 7089, XIX 246.

113 Cf. *Collins*, XXVII 283.1–14. The distinction in the lectures is essentially the same as the later one in the *Metaphysics of Morals*, with the exception that they contain not even a hint that human beings might be the legislators of their own moral laws.

114 See his criticism of heteronomous systems, IV 432.25–433.11 above.

former, the latter is a being without physical needs, and hence without limitations or even subjective principles (maxims).

¶ IV 434.7 Kant now explains that strictly speaking even the formula of autonomy in a kingdom of ends provides a negative criterion like the previous formulations. We are justified in acting on a maxim only if it could be given by an autonomous will as a strictly universal law. The fact that Kant considers autonomy a feature of a perfect will as well as a finite will is a strong indication that we have now reached the territory of pure moral philosophy, or a metaphysics of morals. By comparison, the imperatival character that the moral law possesses for us is quite insignificant.

d. A moral being possesses dignity, not a price

¶ IV 434.20 Kant turns from the description of abstract laws to the discussion of values. He introduces the *idea* of the dignity of a rational being that is robustly independent and self-reliant in practical matters because it 'obeys no law other than that which it itself at the same time gives' (IV 434.29–30).[115] Reason is firmly in charge, unimpressed by the wiles of inclination. Kant is obviously speaking of a creature that lives up to the imperative of autonomy (see IV 439.35–440.13). In his discussion of servility in the *Metaphysics of Morals* he expressly suggests that immoral action divests human beings of their dignity. However, he also suggests that in everyone there is a residue of 'inalienable' dignity that inspires self-respect (VI 434–6; see also VI 462–5). Moreover, reverence is expressly said to be inspired by the *idea* of autonomy at IV 440.10 below.

¶ IV 434.31 In an ideal Kantian commonwealth, *things* are accorded a relative value because they are essentially interchangeable, whereas a *person* possesses absolute value and cannot, morally, be replaced with another person. Compare losing a friend and losing an umbrella. If you lose a friend you do not just shrug your shoulders and get yourself a new one; or not, at any rate, if you behave like a member of a kingdom of ends.[116]

115 Kant has repeatedly emphasised the dignity of practical reason and the command of duty; see IV 405.12, IV 411.2 and IV 425.28.

116 The distinction between price (*Preis, pretium*) and dignity (*Würde, dignitas*) is Stoic in origin. Kant's source is generally believed to be Seneca, *Letters*, 71.33. On the distinction between

¶ IV 434.35 It is tempting to see in 'market price' (*Marktpreis*), 'fancy price' (*Affektionspreis*) and 'inner value' or dignity the values defined by the three types of imperative at IV 414–16. However, from the discussion that follows it would seem that both technical and pragmatic imperatives concern things that have a market value, whereas the fancy price, or price of affection, is a value reserved for things aesthetic.[117]

¶ IV 435.5 Moral action – obeying the laws one imposes upon oneself – opens the gate to the kingdom of ends; and maxims that conform to these laws inspire reverence, independently of whether the actions that result are graced with success. Throughout this paragraph Kant uses qualities like morality, skill, diligence, wit and humour[118] as success terms. Our reactions are provoked by the *exercise* of these faculties, not by their potentiality.[119]

¶ IV 435.29 The foundation of human dignity can be specified further. It is the share we have in the legislation of universal laws and our membership in a possible ideal moral commonwealth. Autonomy grounds the special position of human beings, and of any other rational creature with which we are not yet acquainted.

8 Reflections on the variant formulations of the categorical imperative

a. The connection between the three variants of the categorical imperative

¶ IV 436.8 This brief but crucial paragraph is a translator's nightmare. It contains several linguistic stumbling blocks. First, Kant says that the three ways of representing the principle of morality are at bottom merely so many (three) formulae of that law (*eben desselben Gesetzes*, IV 436.9[120]), i.e. of the single, canonical categorical imperative as stated

things and persons cf. IV 428.18–25. Personhood is linked to the capacity of pure practical reason at IV 438.14–16 below.

117 On the trichotomy see also Kant's *Anthropology*, VII 292, and R 1498, XV 774–81.

118 'Humour' (*Laune*) refers to the amusing talent to take on different moods at will; see *Critique of Judgement*, V 336.1–5.

119 The thesis that 'morality, and humanity in so far as it is capable of morality' (*so fern sie derselben fähig ist*) possesses dignity (IV 435.8) seems to contradict this, but Kant may well be assuming that, as the proverb goes, the proof of the pudding is in the eating: a person possesses dignity in so far as he or she is able to lead – i.e. succeeds in leading – a moral life.

120 'Of the selfsame law', an expression that is equally antiquated and equally ambiguous. It could mean, and has been taken to mean by Kant's translators, that the formulations are

at IV 421.7–8. There is precisely one such principle, but it can be restated in different ways. Kant is warning his readers not to mistake the three variants for principles in their own right. In the lectures, he condemns the pluralism of other ethical theories: 'Where there are already many principles, there are certainly none, for there can only be one true principle' (*Collins*, XXVII 266.8–10).[121]

Secondly, one of these three reformulations naturally and easily (*von selbst*) unites within itself 'the other two' (*die anderen zwei*): the formula of 'autonomy in a kingdom of ends' combines the formal and material features of the previous variants, as indicated in the course of its derivation (IV 431.14). The disputed 'equivalence' of the formulations should not therefore be seen to rest on the claim attributed to Kant by some translators that 'each' or 'any one' of the variants 'unites in it the other two', if only because Kant claims no such thing.[122] The formulations are *equivalent* in that, when applied in a given situation, they lead to the same result. After all, all three variants are, directly or indirectly, derived from the same conception of human volition.

Yet, thirdly, there is a difference between the basic, general formulation of the categorical imperative and the three other formulations.[123] This difference is 'subjectively rather than objectively practical', i.e. they command the same actions but they affect us differently. Each of the three variants employs 'a certain analogy' to bring an idea of reason closer to intuition, i.e. 'a perfect similarity between . . . entirely

three ways of representing one and the same law, which makes good philosophical sense. However, in the light of the contrast implied by the subsequent sentence it is more than likely that the pronoun refers back to the 'principle of morality', the general categorical imperative, that is presented, represented, portrayed (*vorstellen*) in three different ways. *Eben derselbe* is common as a relative pronoun when used on its own; in conjunction with a noun it is ambiguous in the way described above. For the relative construction see e.g. Kant's letter to Theodor Gottlieb von Hippel, dated 24 October 1791 (XI 300, No. 493 [461]). Kant has been invited to luncheon (*zur Mittagsmahlzeit*) by the court chaplain, the Revd Schultz, that coming Wednesday, and he reports that 'Frau Hofpredigerin' has asked him to invite Hippel along 'to the selfsame meal' (*zu eben derselben Mahlzeit*). Are we to suppose that they ate from one and the same plate?

121 This, interestingly, in the context of Stoic law-of-nature theory. A multiplicity of determining grounds were rejected in Kant's 'reply' to Sulzer at IV 411 fn. above. Similarly, in the Methodology of the *Critique of Pure Reason* Kant warns us that in philosophical matters there can be only one proof, and that the dogmatist who enters the philosophical stage with ten proofs certainly has none at all (A 789/B 817).

122 Paton, Ellington and Wood offer the correct first reading, Abbott, Beck, Gregor and Zweig the second. The latter would be a poor argument for the equivalence of the formulations also because it concerns only the variants and not the general formulation, explicitly related to each of the variants in the 'review' at IV 437–40 below.

123 There is a difference between the 'general formula of the categorical imperative' and the variants (see IV 436–7), not between any of the three variants, as the text might seem to suggest, let alone the first and the other two.

dissimilar things' (*Prolegomena*, IV 357.28–9). Analogies do not contribute towards our cognition of an object. They determine only the manner of reflection about the object, which is precisely what the three variations of the categorical imperative do (cf. *Logic*, IX 132). The three notions used in analogy with moral volition are: a consistent, formal 'system of nature', human beings as certain 'ends-in-themselves', and autonomy in a 'kingdom of ends'. This theme is pursued in the 'review', IV 437–40.

¶ IV 436.15 Referring to the *form* of a moral maxim – its universality – Kant uses not the general formula of universal law, but the analogy employed in the law-of-nature formulation (IV 421.18–20). It is obviously considered a variant like the two others; and the variant, rather than the general formulation, is required to yield the third variant as 'inverted teleology', which draws it into the realm of a moral metaphysics. This is frequently overlooked.

¶ IV 436.19 The *matter*[124] of a maxim is its end. That is why the categorical imperative can also be formulated with regard to the ends it commands (see IV 429.10–12). However, the 'ends' to which the second variant refers cannot be ends desired by the agent. They are objective ends, as the limiting condition of all subjective ends.

¶ IV 436.23 The *complete – formal and material – characterisation* of all maxims. Like the three categories of quantity (unity, plurality, totality) that Kant now makes use of, the third variant of the categorical imperative is supposed to arise from a combination of the other two.[125] The formal *unity* of the law, as a command that addresses a *plurality* of ends-in-themselves towards all other rational beings, leads to the

124 The original editions have *Maxime* (maxim) rather than *Materie* (matter), but Arnoldt's conjecture is clearly correct. It does not make sense to say that 'all maxims have a maxim'.

125 For the categories, see *Critique of Pure Reason*, A 80/B 106; for the suggestion that the third category can be interpreted as a synthesis of the other two, see the first 'charming observation' at *Prolegomena* IV 325 fn. (1783), Kant's 1784 correspondence with Johann Schultz (Nos. 208–211 [190–193], X 348–54, and esp. No. 221 [202], X 366–68), and the second edition of the *Critique*, B 110. There is no trace of it in the first edition. In the letter dated 17 February 1784 – Kant was working on the *Groundwork*; see Menzer's academy notes, IV 626–7 – Kant warns Schultz not to regard the combination of the first two categories in each group as a purely mechanical process of blending (*Zusammennehmung*, X 366). The third category is not superfluous; nor is the third variant of the categorical imperative. An act of synthesis is needed in either case. This could count as further proof, if further proof were required, that the three variant formulations of the categorical imperative are not supposed to contain 'each other'.

all-encompassing ideal of a systematic, harmonious commonwealth of all such ends. The law-of-nature formulation serves as the formal variant. It is the vocation of human beings to shape existing nature, as far as possible, in the image of ideal nature. The thesis that the third variant is characterised by a 'complete determination' explains why the 'formula of autonomy' and the cognate 'formula of the kingdom of ends' do not count as two different formulations. The triadic structure also reflects Kant's thesis that form takes the lead by constraining matter.

Kant recommends the 'strict method' of the 'general formula'[126] for practical purposes, but the three variations may be used to secure acceptance for the moral law. They employ vivid analogies and bring the merits of moral action closer to intuition. In particular, the notion of the dignity of a robustly self-reliant 'end-in-itself', which is made fully explicit only in the third, most metaphysical variant,[127] strengthens our reverence for the moral law and thus our desire to lead a moral life. Recall that at the end of Section I the task of a novel 'metaphysics of morals' was said to be the protection and restoration of the natural innocence of our moral faculty by pointing it, in a Socratic manner, to its very own principle (IV 404.4). This has now been achieved. Thus, for Kant, the purpose of his new kind of ethical theory does not just consist in providing a 'decision procedure', even though the 'compass' of the general formulation (IV 404.1) may to some extent fulfil this function. People who have never read the *Groundwork*, let alone the secondary literature, know full well what they ought to do if only they pay due attention to their own moral judgement. Rather, the metaphysics of morals is meant to emphasise the purity, dignity and sublimity of our rational selves, and to encourage us to live up to the better part of our nature in the face of the wiles of inclination and the perils of bad moral philosophy. In substance, it is precisely this stirring metaphysical picture of the vocation of humanity that distinguishes the second chapter of the *Groundwork* from the first. In Section II, Kant moves from the applicable to the metaphysical.[128]

126 When Kant speaks of the 'general formula' he is referring to the first and basic formulation of IV 421, not the formula of autonomy, as Wood suggests (*Kant's Ethical Thought*, p. 188). Kant says that the maxim must be capable of making 'itself at the same time a universal law', i.e. it must be able to avoid contradicting itself in the way described at IV 421–3 above. The general formulation cannot refer to a formula of autonomy because the will – that which, according to Kant, possesses autonomy – is not even mentioned. Moreover, the 'general' formula of the categorical imperative can hardly refer to one of the specific variants.

127 Through the variants of Section II there is thus a 'progression' (*Fortgang*) towards metaphysics.

128 Kant assigns the same role to ethical theory in his essay on *Theory and Practice*; see esp. the Preface and Section I, VIII 275–89.

Fn. IV 436 Kant declares moral philosophy to be a kind of 'inverted' teleology because moral action consciously places a purposeful structure on nature. The analogy is further explored below (IV 438.8–439.34). In Kantian teleology, ends or purposes are a theoretical assumption that help us to understand processes that cannot be understood adequately on the model of efficient causation and its laws; see the teleological second part of the *Critique of Judgement*.

b. Review of the Groundwork *so far: the good will and the formulations of the categorical imperative*

¶ IV 437.5 The analytic project of the *Groundwork* is now complete. Kant summarises the progression towards metaphysics in Section II and connects consecutive formulations of the categorical imperative with the opening statement about the good will, which can at last be defined with philosophical precision.[129] The present paragraph concerns the general formulation and the first variant. A will that is completely good acts only on maxims that it can will as universal laws; or alternatively, to use the analogy of a purely formal system of nature, those that do not contradict themselves as natural laws (see IV 421.7–20). It seems to go without saying that these two formulations of the categorical imperative issue equivalent commands.

¶ IV 437.21 The second variant.[130] Kant paraphrases the line of thought leading up to the formula of humanity as an end-in-itself (IV 427.19–429.13) and relates it to the good will. He defends the peculiar analogy that the subject that contains within itself all other ends should itself be an end, if of a rather peculiar, and superior, kind: an independently existing end, not an end that is to be effected by human action.[131] A good will always takes account of independently existing ends, i.e. subjects similarly endowed with a good will. It never uses them

129 This fact is overlooked by those who like A. R. C. Duncan and J. Freudiger toy with the idea of dismissing IV 421–37 as a rhetorical excursus directed against Garve's *Cicero* (see Freudiger, *Kants Begründung der praktischen Philosophie*, pp. 25–38). Kant *relies* on his earlier findings and *reconnects* them with the concept of moral goodness.

130 The first sentence of this paragraph is easily misunderstood. It is a general statement about a 'nature' (count noun) in the sense employed throughout the latter half of Section II: about an independent being of a certain kind. What distinguishes 'rational natures in general' (*die vernünftige Natur*) from 'other natures' (*den übrigen*) is that they freely elect (moral) ends. This thought is paraphrased at the beginning of the next paragraph. See also my 'Value without Regress. Kant's "Formula of Humanity" Revisited', *European Journal of Philosophy* 14 (2006), 69–93.

131 In this paragraph (IV 437.31), Kant seems to argue that if a subject that contains all other ends is itself an end it must be both more eminent and different in kind.

merely as means but always regards them as fellow ends-in-themselves. Kant emphasises the equivalence of the second variant – now explicitly formulated to refer to maxims – and the general formulation. They are 'at bottom the same' (*im Grunde einerlei*, IV 438.1). The *universal* validity of a maxim ensures that others are never merely contained in it as a means; or, as one might put it, that no-one is *excluded* from my maxim.

¶ IV 438.8 The third variant, as a consequence of the previous two. The dignified status of a finite being is conditional on the fitness of its maxims to be universal laws: it is *successful* autonomous legislation in moral action which distinguishes it from all those beings that inhabit the realm of nature only (*vor allen bloßen Naturwesen*). If the categorical imperative was universally observed, the kingdom of ends would be realised in nature. All human beings would be happy, at least within the limits of their natural constitution. (They would still be vulnerable and mortal.) Yet the prospect of my own happiness in a mere 'kingdom of nature' does nothing to support the authority of the moral law. This 'paradox' is the foundation of human dignity.

Moreover, the permissibility of actions is defined in terms of agreement with the autonomy of the will, another indication that the latter formulation is to be employed just like the general formulation and the first variant, which are expressly set up as permissibility tests. Kant declares the third variant to be equivalent to the general formulation; and he defines a good will – of both the perfect and the duty-bound, finite kind – with one that acts according to self-imposed universal laws.

¶ IV 439.35 We are now in a position to account for the dignity of those who fulfil their duty. It is not due to the fact that they are *subject* to a law. There is nothing particularly sublime about that: quite the contrary. Rather, the dignity of morally good people is grounded in the fact that they are fully in command: they are subject only to a law *that they impose upon themselves*. Reverence is inspired by the idea of autonomy. The reference is to the footnote at IV 401, where reverence – the motive of moral action – was identified with our interest in acting morally.

9 *The autonomy of the moral will*

a. Autonomy and heteronomy

¶ IV 440.16 Autonomyis self-legislation in a twofold sense: first, it is the legislation of a law that originates within the agent's self and,

secondly, a law that he imposes on himself. Returning to the distinction between the legislative and the executive will, we can note that it is the former that imposes the law on the latter, the faculty of choice, which is obligated to obey, but frequently does not. Autonomy cannot be ascribed to either the legislative or the executive will – as the one gives the law to the other – but only to the volitional faculty as a whole.[132] In accordance with what has just been said about the central position of the general formulation, autonomy is defined as 'never to choose except in such a way that the maxims of one's choice are at the same time contained in the same act of volition as a universal law' (see IV 421.7–8). The bindingness of this principle – the fact that 'this practical rule is an imperative' – cannot be established by means of conceptual analysis; it remains a project to be pursued in the last section.[133]

¶ IV 441.3 All rival ethical theories fail because they misjudge the 'source' (IV 441.2) of the principle of morality.[134] They erroneously place it outside the will, and let it be determined by something other than the moral law. (Kant again paraphrases the general formulation of the categorical imperative.) The only motivational alternative to the moral law is inclination, but this would result in heteronomy, i.e. 'other-legislation', the will's being determined by a law that is *not* its own. The argument for this thesis runs as follows: inclination is always directed at the pursuit of an end to be effected. For this purpose, the will must follow hypothetical imperatives of the 'technical' kind, which rely on the regular workings of the natural world around us. When we act on inclination, it is these regularities, which are external to the will, that determine its causality; or, as Kant puts it, it is 'the object, by virtue of its relation to the will', that gives it the law. If the human will was heteronomous, all human behaviour would depend on inclination. We would be incapable of willing an action unconditionally, just for its own sake – at best, we would be administering interests alien to the will. Moreover, when we act immorally we let such natural regularities determine our behaviour. (As we are free, we could of

132 Cf. the puzzle of how the same thing (the will) can be both active (binding) and passive (bound), raised at *Metaphysics of Morals*, VI 417.7–418.23. We need to distinguish two parts of our human existence, *homo phaenomenon* and *homo noumenon*, which are roughly supposed to correspond to the two elements within the human will: *Wille* and *Willkür*.

133 The categorical imperative is a synthetic practical principle because it commands something genuinely new; see IV 420 fn.

134 The project of a metaphysics of morals was already linked to laying bare the 'sources' (*Quellen*) of morality in the Preface (IV 389.1, IV 389.37; cf. IV 392.20). Christian Wolff is criticised for ignoring the different sources of practical determining grounds at IV 391.8–9. See also IV 405.25, IV 407.37 and IV 426.2.

course have chosen to let the moral law determine our actions instead.) Kant reiterates the lesson learnt in Section I that sympathetic action should not be motivated by the alleged value of the happiness of others but rather on moral grounds (see IV 398.8–399.2).

b. Division of ethical theories according to the principle of heteronomy

¶ IV 441.29 Kant claims to have discovered the 'only true path' to the principle of morality. He is alluding to the *Critique of Pure Reason*, which provides the foundation of the metaphysics of nature, just like the *Groundwork* provides the foundation of a metaphysics of morals. On the necessity of a Kantian 'critique' of reason and the destiny of philosophy without it see, for example, the first-edition preface of the former work (A xii). The theme is developed further in the second-edition preface. Note that reason has erred in its theoretical pursuits, not in everyday moral practice.

¶ IV 441.32 In the *Critique of Practical Reason*, Kant assigns empirical principles of heteronomy to Epicurus in the case of 'physical feeling' and to Francis Hutcheson in that of 'moral feeling' (or moral sense). He detects rational heteronomous principles of internal perfection in Christian Wolff and the Stoics, and of the external perfection of God's will in the work of Christian August Crusius (V 40).[135]

¶ IV 442.6 Having established that autonomy is the foundational notion of a metaphysics of morals, Kant returns to the theme of the Preface and introductory remarks to Section II. All empirical principles fail to pay due attention to the non-empirical status of moral commands; egoistic hedonism in particular, in that it reduces moral action to first-personal felicific value and thus destroys the specific difference between moral and non-moral motives. Moral sense theories fare somewhat better. At least they represent an attempt, however insufficient, to account for the value of moral principles and actions in their own right.

Fn. IV 442 Kant indicates that he considers Hutcheson's moral sense theory to be committed to psychological egoism because it implicitly

135 In the *Critique of Practical Reason*, there is, in addition, the category of 'external subjective' determining grounds (as opposed to the internal determining grounds of feeling), occupied by Montaigne's principle of 'education' and Mandeville's principle of a 'civil constitution'. The main dichotomy is that of subjective vs. objective, rather than empirical vs. rational.

or explicitly makes reasons for moral action depend on whether they appear to be agreeable.[136]

¶ IV 443.3 Kant now contrasts the two types of ethical system grounded in reason. The threat of emptiness notwithstanding, an ethic of perfection is said to be preferable to a (purely) theological ethical theory because a concept of God that does not itself refer to his perfection inspires fear in human beings and encourages behaviour that conforms to moral standards from fear and selfish motives, rather than truly moral action from reverence for the law (see the argument at IV 432.25–433.11 above).

¶ IV 443.20 Of the two least criticised theories within the 'divisions', Kant considers an ethic of perfection superior because it preserves the purity of the idea of a good will. It refers moral questions to the court of pure reason, even though this court is incapable of reaching any decisions on empty perfectionist grounds. Perfectionism is a comparatively good foundation for further ethical reflection. In fact, the systems taken up in the first and third variants of the categorical imperative are such rationalist theories.

¶ IV 443.28 There is no need to refute misguided moral theories one by one. It is quite sufficient to know that they are all heteronomous, i.e. fall short of the standards of Kantian autonomy. Even adherents of these theories will by now have been won over.

¶ IV 444.1 Kant reasserts his thesis that a moral, categorical imperative cannot be grounded in any specific object, which can only yield hypothetical imperatives and hence heteronomy (see IV 441.3–7).

¶ IV 444.28 A *good* human will is one that obeys the categorical imperative. This statement officially concludes Sections I and II, i.e. the 'analytic' part of the *Groundwork*.

10 Transition to Section III: How is a synthetic practical proposition possible?

¶ IV 444.35 We now finally turn to the problem of the validity of morality, which has been postponed for so long.[137] In the two 'analytic'

136 Note that 'feeling' (*Gefühl*) refers to our emotional faculty, not to an individual sentiment, type or token; cf. IV 460.3 below.

137 See IV 419.12–420.17, IV 425.7–11, IV 429 fn., IV 431.32–4.

sections, the common notion of morality was revealed to contain a commitment to the autonomy of the will as its metaphysical foundation. But the question of how a categorical imperative is possible as a synthetic a priori practical principle cannot be addressed within the confines of the new metaphysics of morals. The foundation of any kind of metaphysics must be 'critical'. That is why the following third 'transition' moves from the metaphysical notion of autonomy to as much of a 'critique' of pure practical reason as is necessary to complete the task of the present *Groundwork*.[138] It represents Kant's attempt to explain how for human beings autonomy can be more than wishful thinking, a 'phantom of the brain' (*Hirngespinst*, IV 445.8).

BIBLIOGRAPHY

Bittner, Rüdiger. 'Handlungen und Wirkungen', in *Handlungstheorie und Transzendentalphilosophie*, ed. G. Prauss (Klostermann, 1986), 13–26

Foot, Philippa. 'Morality as a System of Hypothetical Imperatives', in *Virtues and Vices* (Blackwell, 1978), 157–73 (first published in 1972)

Gaut, Berys and Kerstein, Samuel. 'The Derivation without the Gap: Rethinking Groundwork I', *Kantian Review* 3 (1999), 18–40

Glasgow, Joshua. 'Expanding the Limits of Universalization: Kant's Duties and Kantian Moral Deliberation', *Canadian Journal of Philosophy* 33 (2003), 23–48

Hill, Thomas E. 'The Hypothetical Imperative', in *Dignity and Practical Reason in Kant's Moral Theory* (Cornell University Press, 1992), 17–37 (first published in 1973)

'Imperfect Duty and Supererogation', in *Dignity and Practical Reason*, 145–75 (first published in 1971)

'The Kantian Conception of Autonomy', in *Dignity and Practical Reason*, 76–96 (first published in 1989)

'Meeting Needs and Doing Favors', in *Human Welfare and Moral Worth. Kantian Perspectives* (Clarendon Press, 2002), 201–43

Höffe, Otfried. 'Kants nichtempirische Verallgemeinerung: zum Rechtsbeispiel des falschen Versprechens', in *Grundlegung zur Metaphysik der Sitten. Ein kooperativer Kommentar*, ed. O. Höffe (Klostermann, 1989), 206–33

Kain, Patrick. 'Self-Legislation in Kant's Moral Philosophy', *Archiv für Geschichte der Philosophie* 86 (2004), 257–306

Korsgaard, Christine. 'Kant's Formula of Humanity', in *Creating the Kingdom of Ends* (Cambridge University Press, 1996), 106–32 (first published in 1986)

'The Normativity of Instrumental Reason', in *Ethics and Practical Reason*, ed. G. Cullity and B. Gaut (Clarendon Press, 1997), 215–54

138 Section III does not render such a critique superfluous, but the remaining elements – such as the representation of the 'unity of reason' – are much less urgent; see IV 391.24–31.

Laberge, Pierre. 'La définition de la volonté comme faculté d'agir selon la représentation des lois', in *Grundlegung zur Metaphysik der Sitten*, ed. O. Höffe, 83–96

Louden, Robert B. 'Go-carts of Judgment. Exemplars in Kantian Moral Education', *Archiv für Geschichte der Philosophie* 74 (1992), 303–22

Ludwig, Bernd. 'Warum es keine hypothetischen Imperative gibt, und warum Kants hypothetisch-gebietende Imperative keine analytischen Sätze sind', in *Aufklärung und Interpretation*, ed. H. F. Klemme, B. Ludwig, M. Plauen and W. Stark (Königshausen & Neumann, 1999), 105–24

McNair, Ted. 'Universal Necessity and Contradictions in Conception', *Kant-Studien* 91 (2000), 25–43

O'Neill, Onora. 'The Power of Example', in *Constructions of Reason* (Cambridge University Press, 1989), 165–86

'Autonomy: The Emperor's New Clothes', *Proceedings of the Aristotelian Society*, Supp. vol. 77 (2003), 1–21

Pogge, Thomas. 'The Categorical Imperative', in *Grundlegung zur Metaphysik der Sitten*, 172–93

Potter, Nelson. 'Duties to Oneself, Motivational Internalism, and Self-Deception in Kant's Ethics', in *Kant's Metaphysics of Morals*, ed. M. Timmons (Oxford University Press, 2002), 371–89

Prauss, Gerold. *Kant über Freiheit als Autonomie* (Klostermann, 1983)

Schroeder, Mark. 'The Hypothetical Imperative?', *Australian Journal of Philosophy* 83 (2005), 357–72

Timmermann, Jens. 'Kantian Duties to the Self, Explained and Defended', *Philosophy* 81 (2006), 505–30

'Value Without Regress. Kant's "Formula of Humanity" Revisited', *European Journal of Philosophy* 14 (2006), 69–93

Section III: Transition from the metaphysics of morals to the critique of pure practical reason

The third section of the *Groundwork* begins with the metaphysical concept of autonomy, which concluded the analysis of the second. However, Kant now ventures beyond explication: he tries to prove that our intuitive trust in a categorical command as the supreme normative principle of human conduct is justified. At the outset of Section III, the odds are heavily stacked against morality. Mere analysis cannot establish the validity of moral concepts; nor can they be justified with recourse to experience. For all we know about human agency so far we may well doubt whether moral terms have any significance for ourselves or for others at all. Autonomy might be an empty concept, and morality an illusion. Section III now asks whether we can make proper sense of autonomy. How is a categorical imperative possible, as a synthetic practical proposition a priori? A successful answer to this question is the element of a 'Critique' of pure practical reason required to complete the foundation of a metaphysics of morals.

The project of Section III divides into three stages. First, by way of preparation, Kant tries to trace the origins of the moral law. If autonomy is a legitimate concept, we have shown that human beings are *subject* to the moral law. Moral self-legislation is revealed to be a consequence of possessing freedom of the will; and we are conscious of our freedom – if indeed we are free – because we are members not just of the familiar world of experience but also of an intellectual or 'intelligible' world. We now understand the 'whence' of the moral law. To make this point, Kant leads us into a vicious circle that we escape only when, from a practical point of view, we realise the inevitability of this assumption. We come to see ourselves as members of two spheres of legislation: natural (heteronomy) and rational (autonomy). Secondly, Kant must show how the law of autonomy is *connected* with the finite human will. This is the core of the deduction of the categorical imperative. The assumption that we are subject to two radically different kinds of law does not yet explain the possibility of a categorical imperative as a synthetic practical principle. The intelligible world is shown to contain laws that are authoritative with regard to the world of sense.

The third task of Kant's justificatory project would consist in explaining the *motivating* or *necessitating* character morality possesses for us,

i.e. the interest we feel we must take in doing our duty. (The possibility of acting on *any* law depends on a corresponding interest.) But this cannot be done. We have reached the extreme boundary of moral philosophy. The deduction is graced with limited success. We can open up a philosophically respectable space for morality, but then we have to concede defeat: the human mind can only comprehend the incomprehensibility of an unconditional practical law. It is no accident that in the course of answering the first two questions Kant feels he has to return to the views commonly held by human beings, which formed his starting point at the outset of Section I. In particular, he needs common moral consciousness to confirm that autonomous volition is unconditionally good. The perilous nature of the present deduction means that pure moral philosophy cannot get away from common rational moral cognition. In the end, the deduction is just an inference to the best explanation, if one that pure practical reason compels us to make.

1 The concept of freedom is the key to the explanation of the autonomy of the will

¶ IV 446.7 The third section, like the previous two, begins with a statement about the will. Kant is trying to explain the nature of autonomy, the defining property of a moral will introduced at the end of Section II. We have a clear conception of what autonomy means, but would like to be reassured that human beings are endowed with a will that has this rather fantastic property. That is why Kant steps back to demonstrate that freedom is the 'key' to explaining the autonomy of the will (IV 446.6). A will is autonomous *by virtue of* being radically free and independent of external determination. Freedom is the *ratio essendi* of moral self-legislation: it is *because* we are free that we must impose on ourselves our very own rational law (see second *Critique*, V 4 fn.). Freedom is not merely a necessary but even a sufficient condition of autonomous moral agency. Oddly enough, 'can' implies 'ought'.[1]

Let us turn to the details of the argument. Kant considers the will a kind of 'causality' because it is the power to produce effects, in the world of experience. Only a rational being possesses a will. (Only a rational

1 The radical negative freedom of a finite will implies that it is subject to the categorical imperative. However, the justified conviction that a will is free falls short of explaining *how* a categorical imperative is *possible* as a synthetic practical principle, which is why the actual deduction of the fourth subsection is needed even once our self-conception as free agents has been corroborated in the third.

being can act according to the representation of objectively valid laws; see IV 412.26–8). The freedom of such a will consists, first and foremost, in the power to ignore external – and hence 'alien' (IV 446.9), i.e. natural or possibly divine – influences like human inclinations. Freedom is not opposed to determination as such, but to the wrong kind of determination. Kant assumes that the will is radically independent of natural influences, a property called 'transcendental freedom' elsewhere.[2] By contrast, non-rational beings like animals are determined by the causal influences of nature. They lack autonomy, and their behaviour can be subject to external conditioning and manipulation.[3]

As in the Preface (IV 387.14–15), Kant conceives of 'laws of freedom' and 'laws of nature' as parallel descriptive laws that govern separate realms. They concern two kinds of metaphysical doctrine. Freedom and natural necessity are qualities of types of causality that operate in different causal spheres. This idea makes sense only on the assumption that we are at present discussing the workings of a pure will in the abstract – the will that is the object of a pure moral philosophy or 'metaphysics of morals' – not a complete human will, which can go astray because it is exposed to conflicting motivational forces.[4] Moreover, the realm that is governed by the law of freedom has not yet been officially introduced.

2 See *Critique of Practical Reason*, V 97.1. Neither comparative psychological freedom of action nor independence merely of unwelcome inclinations will do. (There can be no inclination towards morality; and prior to rational assessment all inclinations are on a par; see IV 398.15.) However, readers of the *Groundwork* are not yet familiar with the doctrine of transcendental idealism. This is significant. There is a veiled reference to the first *Critique* at IV 448 fn., but the decisive distinction between things as they are in themselves and as they appear to us is introduced, with a flourish, only as our last hope to escape the vicious circle at IV 450.30–4.

3 Kant envisages a rather stark division between free, rational – and ultimately moral – beings that possess a will, and animals, which lack a will and function according to mechanistic laws. There seems to be little room for rational beings with a lesser will, in which the role of practical reason is subservient to inclination, or no will at all. This is surprising, because the fear that we might have a will that is subject to hypothetical imperatives only pervades Section II (e.g. IV 419.16–19). Kant presumably thinks that in so far as a will is rational (IV 446.7) it is free. In a lesser kind of will reason would not by itself be practical; it would never determine the law of action but always borrow it from nature. If reason is strictly equated with practical reason, a lesser volitional faculty would not be a will. See also K. Ameriks, 'Kant's *Groundwork* III Argument Reconsidered', in *Interpreting Kant's Critiques*, pp. 237–8.

4 This problem was prominently raised by H. J. Paton. He complains that the analogy between the law of nature and the moral law is flawed because of the descriptive nature of the former and the prescriptive (imperatival) nature of the latter (*The Categorical Imperative*, p. 211). But the moral law is descriptive of the actions of a pure will just as the causal laws of nature describe the series of events in the world. Human beings also possess a will that is at home in the realm of freedom. Problems arise because there is more to human volition. See IV 454.6–19. The theme of the two parallel spheres of legislation is continued at IV 452.22 and IV 453.17. Kant is also alluding to the need to reconcile natural necessity and freedom with regard to beings that are subject to both kinds of causal law; see IV 455.11–457.3.

¶ IV 446.13 So far, the only thing we have learnt about free will is what it is not – which, if correct, is uninformative. Fortunately, a *positive* conception of freedom as autonomy 'flows from' (*fließt aus*, IV 446.14) the negative notion of freedom as complete independence of natural determination.[5] What does Kant's argument look like?

Let us accept the thesis that the will is that by virtue of which rational beings are causally efficacious. Kant now introduces the additional premise that the concept of causality analytically implies that causes and effects are conjoined by unvarying laws.[6] It follows that if a will effects changes it must be in accordance with causal regularities. However, these laws cannot be the familiar causal laws of nature. The previous paragraph revealed that natural influences, by definition, cannot govern the workings of a free will as such. The law of its causality would then be taken from nature. But what can possibly determine the causality of freedom? A 'lawless will' is a conceptual impossibility[7] – a will must always be determined to act by *some* law. As non-natural causal laws external to the will are not in sight, only one possibility remains: that a free will gives itself a law; and this is *autonomy*. In sum: a will possesses autonomy by virtue of its transcendental freedom.[8] 'Heteronomy' means that the cause of an event must in turn be determined by some other, prior cause. But if freedom is *not* heteronomy, it must be 'autonomy': a will can be active just on its own. The disjunction ἕτερος and αὐτός leaves no third possibility.

Autonomy – the property of a will to be a law to itself (IV 447.2–3) – is then equated with the general formulation of the categorical imperative: the principle only to act on a maxim that 'can also have itself as a universal law as its object' (IV 447.4–5; see IV 421.7–8). Kant concludes that, as this is the formula of the categorical imperative and

5 On negative vs. positive freedom see also *Critique of Practical Reason*, V 33.15–18.

6 This is a Kantian commonplace. See *Prolegomena*, IV 257.25–7, *Critique of Pure Reason*, B 5 and A 539/B 567.

7 *Unding*, IV 446.21. Independence as such does not suffice to *determine* the will to act. Negative freedom cannot be a 'determining ground' of the will, i.e. that which gives it a causal law. Moreover, mere irregularity raises the question of why we should impute such actions to agents at all (see R 3860, XVII 316). Some regularity other than that of natural causality must therefore govern free actions. This is a recurring theme in Kant's moral philosophy. Kant explicitly rejects the equation of freedom of choice and freedom of will. Rather, human freedom of choice between good and evil is an expression of our weakness, the limitations of our faculty of volition. And yet, it seems that the finite human will, which faces a choice of *different* laws, does – problematically – not have a single law that determines its causality. See Appendix E.

8 The argument, if intelligible, rests on a number of controversial assumptions. Kant's notions of freedom, causality and their respective laws can be challenged. He does not provide further arguments to support these claims.

the principle of morality, 'a free will and a will under moral laws' are identical (IV 447.7).

Two notes. First, Kant does not need to offer an argument for the equivalence of autonomy and morality at this stage. He is – rather too hastily – relying on the results of the previous, metaphysical section (see, for example, IV 436.8–10, IV 440.14–32). Secondly, the equation of autonomy and morality glosses over different kinds of will. Mentioning the categorical imperative at IV 447.6, Kant returns from the rarefied purity of moral metaphysics to a mixed or finite will. However, as the subsequent paragraph shows (IV 447.8–14), Kant is not yet concerned with the synthetic or imperatival character that the moral law possesses for human beings – which may explain his reference to the 'formula' of the categorical imperative (IV 447.5), rather than the categorical imperative itself. The expression 'a will under moral laws' is deliberately general. It covers a mixed will, the isolated pure will and a perfect or holy will alike.

¶ IV 447.8 Morality and its principle follow *analytically* from the presupposition of freedom of the will. Yet the supreme principle of human conduct in the shape of the categorical imperative, which commands actions because they follow from the moral law and identifies them as morally good, is nonetheless a *synthetic* practical principle;[9] and it is the moral law in the guise of a synthetic practical principle that Kant intends to establish by means of his deduction.

How are we to make sense of the analytic/synthetic contrast at this point in Section III? Three notes. First, recall the hiatus in the argument of Section I. Kant started with an explication and preliminary defence of the absolute value of the good will (IV 393–6) but quickly switched to an analysis of duty (IV 397.1–10), which is defined in terms of acting from reverence for the law, not of doing good (IV 400.18–19). Consequently, the first statement of the categorical imperative at IV 402.7–9 is not directly derived from an analysis of good volition.[10] Secondly, and relatedly, there are no imperatives for a good will as such. The concept of duty contains that of the good will, but not conversely. Action for

9 If Kant's 'the latter' (*das letztere*, IV 447.10, suppressed by most translators) is referring back to the 'principle', he is guilty of a minor equivocation because in the summary 'its principle' (IV 447.9) is the principle of autonomy, which need not take the shape of an imperative, whereas the principle that is synthetic is imperatival.

10 The categorical imperative is a synthetic practical proposition analytically derived from the concept of duty: a more precise statement of the equally practically synthetic principle that human beings *ought to do their duty*. It tells us what precisely our duty is.

the sake of the law is identified with morally good action, but not every conceivable good act is done from duty. Whether it is depends on the nature of the will in question. It is the defining feature of the necessitation of a finite will by a synthetic practical principle a priori that in morally good action it does not act to bring about some perceived good. It acts for the sake of 'the law' (general formulation, first variant) or for the sake of some 'objective' end (second variant). Such action *constitutes* goodness. Acting *for the sake of* the good would be a mark of heteronomy.[11] Thirdly, the preceding argument from freedom to autonomy fails to mention any normative moral concepts. The notion of autonomy was used in the thin, literal sense of 'self-legislation'. Kant now formulates the synthetic principle of morality in terms of what a *good* will would do, the ideal to which human beings must aspire in their conduct[12] – though not by pursuing the good. It is precisely a principle of morally good action – as distinct from action for the sake of something that is perceived as good – that requires corroboration. The principle of morality as deduced in the fourth subsection at IV 453–5 is: an absolutely good human will accords with the principle of autonomy (IV 447.10–12).[13]

The categorical imperative is 'synthetic' because corresponding volition and action do not follow from our natural desires. It does not, like hypothetical imperatives, 'analytically' command that we adopt a law that enables us to do what we would like to do. This raises the question of how a categorical command of reason is possible: how the law of autonomy can determine *our* will. Thus the project of Section III would not be complete if our conviction that we are free was confirmed beyond reasonable doubt in the second subsection (which it is not). Admittedly, on the basis of the analytic argument from transcendental freedom to autonomy we would be in a position to conclude that we 'stand under' moral laws. But that in itself would leave the syntheticity of the categorical imperative unexplained. For that purpose, Kant must establish a connection – completely unobvious so far – between the law of freedom and a human will that recognises the one

11 See Kant's rejection of H. A. Pistorius' criticism that the definition of moral value ought to be prior to the principle of morality in the second *Critique*, V 8.25–9.3; See also V 62.36–64.5 and his ruthless reinterpretation of the 'old formula of the schools', V 59.12–13.

12 Again, Kant is drawing on the results of Section II; see the summaries at IV 437–40.

13 This sentence encapsulates the results of Sections I and II in a nutshell; see the summary at IV 437.6–7. The 'absolutely good will' (*ein schlechterdings guter Wille*) at IV 447.11–12 is thus the good *human* will. As D. Schönecker has shown, it is this synthetic principle that must be deduced (see *Kant: Grundlegung III*, pp. 154–6).

kind of law as being more authoritative than the other – a problem that does not arise with regard to a will that is completely free and governed solely by the law of reason, or for a pure will viewed in isolation. It is only when this connection is made that the commonsense belief in the absolute goodness of moral action is vindicated (see IV 455.4).

Kant now argues that as in the case of theoretical synthetic judgements a priori some third representation[14] is needed in order to show how a practical synthetic proposition a priori is possible. The positive conception of freedom (autonomy) 'provides' (*schafft*, IV 447.17) or 'points to' (*weiset*, IV 447.21) this third element. Kant's choice of words at IV 447.18–19 contains a hint as to what it might be. The 'third something' does not, as in the case of natural causes, relate to the nature of the world of sensation. It is rather the idea of another kind of causal power: a pure will, located in a superior intellectual world with its very own laws (IV 454.12–14). Unfortunately, we are not yet in a position to appreciate this point. Kant first rather mischievously leads us into a vicious circle (IV 450.18–23) in order to reveal the new world to us on our way out. The 'preparatory argument'[15] therefore takes place in the subsequent second *and* third subsections. The deduction[16] of the categorical imperative follows after our lucky escape, in the fourth.

14 For a synthetic proposition two cognitions (*Erkenntnisse*) must be combined 'with a third' (*mit einem dritten*, IV 447.16; see IV 447.18), but Kant is not speaking of a 'third cognition', as Abbott, Beck, Ellington, Gregor, Zweig and Denis have it, despite his frequent use of *Erkenntnis* as a neuter noun. What is needed is 'a third something', in fact a third representation. Kant, the copyist or the typesetter could have avoided this misunderstanding by using an initial capital letter (*mit einem Dritten*/*dieses Dritte*) to flag substantival use, as required by more regular modern spelling conventions. (Some of his later editors, such as Vorländer, have done it for him.) In the first *Critique*'s chapter on the subject, Kant speaks of *ein Drittes* (capital D, without *Erkenntnis*) throughout. The 'third something' is experience in the case of a posteriori judgement; in that of synthetic a priori judgements it is – minimally – time (A 154–8/B 193–7).

15 The term (cf. *Vorbereitung*, IV 447.25) was coined by H. E. Allison, who erroneously confined it to the following second subsection (*Kant's Theory of Freedom* (Cambridge University Press, 1990), pp. 214–18). This interpretation is now widely accepted (see e.g. T. E. Hill, 'Editorial Material', pp. 132 and 135). However, as we shall soon see, no explicitly non-moral evidence for human freedom is given in the second subsection, and the preparation of the deduction is still incomplete. In fact, because Kant obscures the synthetic nature of our consciousness of freedom, the second subsection leads us straight into the vicious circle (see Hill, 'Editorial Material', p. 139, for a similar suggestion).

16 Why does Kant speak of the 'deduction of the concept of freedom' at the end of this paragraph (IV 447.22–3), rather than a deduction of the highest principle of morality (see IV 463.21–2) or the categorical imperative? Perhaps because freedom has been revealed as the ground of morality; see the 'entitlement' to freedom (*Rechtsanspruch*) at IV 457.4–5; or because the deduction takes the shape of justifying the application of the – positive – conception of freedom to ourselves, i.e. autonomy.

2 Freedom as property of the will of all rational beings

¶ IV 447.28 We have learnt that the validity of moral laws follows from the assumption of radical negative freedom. Consequently, if we had proof that the human will was free we would be entitled to infer that we are subject to moral legislation – but would not yet be in a position to understand its synthetic nature or its psychological force.

The second subsection fulfils two modest tasks. First, it addresses the question of what at the present stage counts in favour of the assumption of freedom. Kant answers that freedom must be presupposed as the property of the will of *all* rational beings, not just the human will or just one's own. Importantly, the reason is that universal and necessary moral laws must be binding for us qua rational beings. This assumption has been in place throughout the argument of the *Groundwork*, ever since it was established in the Preface at IV 389.11–23. It is restated at IV 447.30–1. Kant thus infers, on the basis of the necessary character of moral commands, that morality must be binding, for ourselves and others, by virtue of our possessing rational faculties; then, on the grounds that we must attribute spontaneity (*Selbsttätigkeit*) to reason – which is incompatible with its being influenced by the senses; see, for example, R 4220, XVII 462–3 – that the will of a rational being is free. For practical – i.e. moral – purposes (*in praktischer Absicht*, IV 448.21) we must conceive of ourselves and all other rational beings as free. There is a strong hint that, to use the Latin terminology of the second *Critique*, morality is the *ratio cognoscendi* of freedom (V 4 fn.).

Kant, secondly, defends the thesis that a being that must by necessity think of itself as free *is* free for all practical purposes, even if freedom cannot be known, explained or proven by theoretical reason, a claim which is intelligible only on the background of the solution of the antinomy of freedom in the *Critique of Pure Reason*. Appearances notwithstanding, the threat of natural determinism does not rule out human freedom.

This subsection is no more than a brief reminder. We are still unacquainted with the reason why we – and possibly other rational creatures – act under the idea of freedom. Kant merely confirms that we do.[17]

17 The grounds of our assumption are obvious only once the circle (IV 450.18) has been removed: we consider ourselves free as members of an intelligible world (see particularly IV 452.31–5). If, as has often been argued, the present subsection is taken to provide conclusive evidence of human freedom, and freedom is the *ratio essendi* of autonomy, the deduction

Fn. IV 448 The aim of the footnote is similarly modest. Alluding to the discussion of the first *Critique*'s third antinomy, Kant asserts that freedom cannot be proven or explained – or indeed rejected – by scientific or theoretical modes of investigation (see A 532/B 560 ff.). It is an idea, and as such incapable of being either confirmed or rejected by experience. As theoretical philosophy is silent, it is legitimate to argue for freedom from a practical point of view. The boundaries of what a deduction can achieve are thus quite narrow right from the start (see IV 455.10). The problem of human free will shares its critical destiny with the other two subjects of pre-Kantian metaphysics: the indestructibility of the human soul and the existence of God.

3 The interest attaching to the ideas of morality

a. Preparation of the 'circle': our consciousness of freedom and morality are not grounded in any conventional interest

¶ IV 448.25 Kant for the first time mentions the suspicion that his argument for human autonomy might contain a vicious circle. Morality was 'traced back' to the idea of freedom in the first subsection of this section, and in the previous subsection it was argued that freedom needs to be presupposed if we are to think of ourselves and other creatures like us as rational agents bound by moral laws. We have not yet been presented with an independent grounding of the idea of freedom (*diese*, IV 448.26).

¶ IV 449.7 So what, if anything, connects us with the 'ideas of morality', such as freedom and autonomy? Where do they come from? Why do they concern, touch or apply to us at all? We are still in the dark about that. However, one candidate can immediately be ruled out: it is not a conventional 'pathological' interest, which makes us act in accordance with the laws contained in a hypothetical, not the categorical

of the moral law might seem to come to a premature end at IV 448.22 and the remaining fifteen pages of Section III would be superfluous. But this line of argument is mistaken. Not only would the task of the deduction still not be complete because we would not yet understand how the law of autonomy is connected with the human will, there is at present no evidence for our trust in our free will other than the necessity and universality of the moral law. Kant has just *closed* the circle.

imperative.[18] The moral law is directed not at an object that we would like to bring about but solely at the willing of the action itself.[19]

The precise nature of the question at IV 449.11–13 is vital for our understanding of Kant's justificatory project as a whole. *Why*, Kant asks, must I subject myself to the principle of morality and do so simply as a rational being? There are two possible readings:

(i) Why is it the case that, by virtue of being a rational being, I am subject to the moral law?
(ii) What reason is there for me to subject myself, as a rational being, to the moral law?

The *first* question is essentially explanatory. It is that of a morally decent person whose trust in the supreme authority of ethical commands is challenged by the elusiveness of their source as well as the obvious threat of natural determinism (see IV 453.9). This person is requesting a broadly *metaphysical* reason for the authority of a law that he accepts, or would like to accept. To use Kant's rather paradoxical expressions: why do we feel compelled *to take an interest* in acting for the sake of a law *that does not really interest us*?[20] The *second* question is that of a radical moral sceptic who, say in the face of robust self-regarding interest, asks for a *normative* reason why he should take up the moral point of view at all. A persuasive answer to the second question is also likely to satisfy those who are worried by the first, but not vice versa.

In the final section of the *Groundwork*, Kant is addressing the first question, not the second. We would like to know *why, not whether* moral commands are valid; and this is done by showing *how* the categorical imperative, as a principle of autonomy, applies to the human will. Kant

18 It is important to note that the possibility of adopting *any* law in one's maxim – and hence of any action – depends on the presence of a suitable interest. Even a free agent cannot just act in any conceivable way; he or she must, directly or indirectly, have an interest in a specific action and its corresponding maxim. (That is the reason why maxims cannot be modified 'at will' to pass the test of the categorical imperative.) If so, there must be some specifically moral interest; and it is this interest – reverence for the moral law – that proves elusive throughout the *Groundwork*; see IV 459–60. For the distinction between two kinds of interest see IV 413–14 fn. and IV 459–60 fn.; for the reliance of maxims and hence actions on some interest see *Critique of Practical Reason*, V 79.24–5, V 90.36.

19 This was the upshot of Section I (see IV 412–13 fn.) and explicitly stated at IV 432.5–24; see also IV 425.32–4, IV 459–60 fn. and IV 460.24–5.

20 See Kant's description of the sceptical philosopher who does not challenge the correctness of moral views but merely regrets that we lack the capacity to act on them (IV 406.14–25). For a similarly paradoxical description of moral interest see *Critique of Judgement*, V 205 fn.

pursues the project of justification by explanation (in the broad sense of giving a plausible account). This involves tracing the source or origin of a law the status of which is uncertain because it is not grounded in any interest that must be presupposed. In the end, the worried moralist will rest reassured if he succeeds in understanding, as far as possible, the nature of morality (see IV 461.9–14). However, if no convincing answer can be given to the first question he will despair and defect to the amoralist camp.[21]

Hypothetical imperatives are commands that can be justified with reference to the fact that they apply only to those of us who take an interest in realising the end in question.[22] We subject ourselves to a specific law only if it serves to realise an end we desire to bring about. But the question why we subject ourselves to the moral law cannot be answered with reference to some such specific interest. The moral law could not otherwise apply unconditionally and universally. (This is precisely what distinguishes the categorical imperative: that it is valid without any presupposition for all human beings qua rational creatures.) We do not act for the sake of some given end that we wish to pursue. No interest *impels* us. However, as rational beings, Kant argues, we necessarily *take* an interest in the law, and feel obligated to act accordingly.[23]

¶ IV 449.24 So far, our moral consciousness appears to be the only reason why we must attribute to ourselves a free will. If so, freedom cannot in turn be used to ground the special value of moral actions (IV 449.34) or moral agents (IV 449.36). (The vicious circle is officially formulated two paragraphs further down.) In fact, goodness does not feature in the arguments revolving around freedom and autonomy at

21 Kant does not take the traditional amoralist's question of why we should be moral at all seriously. If it is thought to call for some antecedent end or interest, it is explicitly rejected in the concluding remarks at IV 463.11–25. We cannot give reasons why every rational being should obey a command that is, after all, unconditional. Moreover, it is doubtful whether anyone would be in a position to persuade a radical amoralist. Someone who opposes, rejects or altogether fails to recognise the commands of pure practical reason is unlikely to be swayed by rational argument. He would be a case at best for the psychiatrist, at worst for the prison authorities, but not for the philosopher. One might try to expose him to shining examples of virtue to elicit his admiration, such as the case of the hardened villain at IV 454.21–2.

22 Note that instrumental justification is merely preliminary. It is a necessary condition of the rationality of pursuing some end, but it is not sufficient since the end must be rational for a rule to become a command of reason.

23 Kant returns to the idea that a moral ought is the expression of a pure rational will under subjectively difficult conditions (IV 449.16–23) in the course of the main deduction at IV 453.19–31 and IV 454.6–19 below.

all (see IV 447.10–14). If this was the end of the matter, Kant argues, the *Groundwork* would be a conceptual game of limited philosophical value. We would not even be in a position to dispel the worries of those sympathetic to the moral cause.

¶ IV 450.3 In an excursus, Kant dismisses another candidate that might be put forward as the source of moral goodness: the interest in being a good person. He claims that over and above the obvious interest in our own happiness[24] we do indeed take an independent interest in our *worthiness* to be happy (see also the second *Critique*, V 130.6–10). Is that the reason why the moral law obligates us? Kant's reply is a resounding No. Our interest in being worthy to be happy is itself just an expression of our consciousness of the supreme value of morality; and it is the nature of *that* value that we are struggling to understand.

b. The suspicion of a 'circle': freedom and morality

¶ IV 450.18 Kant now makes explicit the worry that so far has been mentioned only in passing. Our account of the bindingness of moral norms appears to be caught up in a circle.[25] We jump from one proposition to the other, and back again, and cannot point to an independent grounding of morality – in the words of the previous paragraph, we still do not know *whence*[26] the moral law obligates us. Existence cannot be proven by means of analytic arguments; and, to put it mildly, there is no evidence of the authority of duty, autonomy or freedom in the realm of experience. We must go beyond explication to account for the possibility of a synthetic practical principle, but the foundation on the basis of which this can be done is not yet in sight.

Kant's actual formulation of the vicious circle at IV 450.19–23 is hardly a paradigm of philosophical clarity. The first half of the sentence in particular has puzzled readers for generations. However, the preceding paragraphs as well as the summary of the resolution (IV 453.3–11)

24 On Kant's formal definition of happiness this is true almost by definition; see IV 415.28–33. In a world like ours worthiness to be happy does not by itself make the agent happy.

25 It is 'a kind of circle' (*eine Art von Zirkel*, IV 450.18) because the argument from morality to freedom contains a hidden synthetic move that allows us to escape the circle after all, not because there is no genuine *circulus in probando* in the first place. (For the latter view see Schönecker, *Kant: Grundlegung III*, pp. 317–96.) Kant similarly accuses perfectionism of circular reasoning – of presupposing that which is to be explained: morality – at IV 443.4–10.

26 *Woher*, IV 450.16. 'On what grounds' (Beck, Gregor, Zweig, Denis) or 'how' (Paton, Ellington) rather obscure Kant's clever quasi-spatial expression. Only Abbott and Wood have the literal 'whence'.

suggest the following reconstruction. The first half does indeed contain the inference from morality to freedom: it was the second subsection above that gave rise to this suspicion (IV 447–8). Morality implies universality and necessity, which can only be grounded in reason; but reason must be independent of alien influences, which is (negative) freedom. In short:

Query: Why do we consider ourselves as free?
Answer: Because we think of ourselves as subject to moral laws.

The hypothetical Sections I and II rely on this assumption, even if freedom does not feature prominently in them.[27]

Less controversially, the second half of Kant's sentence contains a reference to the analytic argument at the outset of Section III (IV 446–7). On the presupposition of freedom we return to the validity of moral laws:

Query: Why do we think we are subject to moral laws?
Answer: Because we attribute to ourselves a free will.

We would have made significant progress towards corroborating the authority of the categorical imperative if the second question had received a persuasive answer, but this is obviously not the case. Those who enquire into the credentials of moral imperatives receive no satisfactory answer when they are told that they are subject to the moral law because they are free even if they accept both the inference from freedom to autonomy and the identification of morality and autonomy. Morality cannot be grounded in freedom if freedom is in turn grounded in morality. There is still no indication of the reason *why* it is the case that we must conceive of ourselves as free and rational beings other than the argument that takes the moral law as its starting point. Freedom of will and moral self-legislation are revealed to be 'reciprocal concepts' (*Wechselbegriffe*). They have the same sphere (see Jaesche's *Logic*, IX 98.5–6). Every will that is free stands under its own moral laws; and every will that stands under its own moral laws is free.[28]

27 This is one possible explanation of why Kant says that we 'afterwards' (*nachher*, IV 450.21) pass through the other half of the circle. However, the argument from freedom to the moral law also comes second in an overall argument for the validity of morality. Kant is primarily trying to show how it is that moral laws apply to us, not that we are free – see the summary at IV 453.3–11.

28 The circle is more apparent if one inserts the mediating concept of autonomy. Strictly speaking Kant should have distinguished four concepts that share the same sphere: negative freedom of the will from natural determining causes; positive freedom of the will to be itself

This is philosophical progress, of sorts; but it does not help us to dispel the worries of a well-meaning sceptic who struggles to understand the phenomenon of morality.

c. The escape: we step outside the circle when we consider ourselves members of an intellectual world

¶ IV 450.30 Fortunately, there is an opportunity to escape[29] the circle that has not as yet been explored. So far, the consciousness of our freedom seemed to follow from our conviction that we are subject to moral laws; and the argument for freedom, via the independence of a rational will, was thought to be analytic. Kant now asks us to entertain the thought – soon to be explained in some detail – that when in deliberation we view ourselves as free we do so from an altogether different standpoint (*Standpunkt*).[30] Kant does not explicitly say how exactly this helps us to evade the circle; but he presumably thinks that the argument from morality to freedom, even if occasioned by moral consciousness, is revealed to contain a hidden synthetic proposition after all. It is the judgement that 'I am free' that enables us to step out of the analytic circle.

This strategy is reminiscent of the 'Copernican revolution', as described in the second edition of *Critique of Pure Reason* (B xvi). In his

active: spontaneity; the causal capacity of the will to act in accordance with its own laws: autonomy in the literal sense of the word; and the capacity of the will to perform moral actions.

29 *Auskunft*, IV 450.30; see Kant's worry that we may not find a way out of the circle (*herauszukommen*) at IV 450.19. (Incidentally, these expressions would hardly make sense if the circle was not meant to be a genuine circle from A to B and back from B to A.) This is a paradigm case of how Kant's eighteenth-century German can mislead his modern readers, who naturally assume that he is referring to some piece of information (e.g. *Telephonauskunft* = telephone directory enquiries), not an escape route or remedy. The modern equivalent of *Auskunft* in this sense of the word is *Ausweg*. Wood ('way out') and Zweig ('route') get it right; all other translations rather obscure the point: 'shift' (Paton), 'resource' (Abbott, Gregor), 'recourse' (Beck, Ellington).

30 See IV 425.33, where the *Standpunkt* or foundation of morality was said to be uncertain. Kant reaffirms the doctrine of the two standpoints in the second *Critique* (V 97.32–98.5). It presupposes his transcendental idealism as exposed in the first *Critique* (cf. B xxvi) but it should not be confused with, or cited as evidence for, the modest epistemological 'two-aspect' reading of Kant's idealism popularised by G. Prauss and H. E. Allison, according to which thing-in-itself and appearance merely represent two different perspectives on one and the same thing. (There is a hint of this in Hill, 'Editorial Material', pp. 102, 106 and 143.) Kant is not concerned with viewing one and the same thing (a human being) from two different points of view. Rather, he encourages us to entertain the possibility that for practical purposes one and the same being can *take up two different positions*. The doctrine of the two vantage points and the introduction of a world of intelligences supports a reading that, metaphysically, is anything but innocent. It assumes that the one world can *determine* the shape of the other (but not vice versa).

theoretical philosophy, Kant does not have a knockdown argument for the thesis of transcendental idealism that objects must conform to our cognition. He suggests that it might be worth exploring whether the discipline of metaphysics can become a proper science on the assumption that they do. Of course, as human reason takes an inevitable interest in this branch of philosophy a constructive proposal as to how to turn metaphysics – cognition of things a priori – into a proper science is very welcome indeed. Similarly, Kant does not have a conclusive argument to prove his thesis that we are members of a world radically different from, and superior to, the realm of natural causes. He merely suggests that the justification of morality, which we consider necessary, can be saved only when we come to realise the philosophical attractions of this assumption.

However, there is one notable difference. The doctrine of the two standpoints reflects the way human beings intuitively conceive of themselves as agents. Our consciousness of the authority of necessary and universal moral laws almost literally propels (*versetzt*) us out of the world of sensibility. By contrast, Kant commonly suggests that the basic assumption of his critical metaphysics of nature takes us away from pre-philosophical human understanding (but see the following paragraph).

¶ IV 450.35 Kant now begins to make good his promise to return to the views of common, pre-philosophical understanding (IV 392.21).[31] To illustrate the point of view that opens up the way out of the circle, he refers to the naïve belief in some hidden dimension that grounds empirically observable things, and a human soul that does not belong to the material world. He attributes to 'the commonest understanding' (IV 450.37) the proto-critical belief that knowledge is confined to 'appearances' that are grounded in unknowable 'things in themselves' (IV 451.7–8), then – following up on that – an admittedly 'raw distinction between a *world of sense* and a *world of understanding*' (IV 451.18). It is unclear whether Kant thinks this distinction is motivated by broadly practical concerns or not.

The first *Critique* distinguishes between things as they appear to us and things as they are in themselves, independent of the limitations of the human mind. Knowledge of things is confined to the empirical realm

31 He does so again when he completes the 'deduction' at IV 454.20 below.

of appearances. Theoretical reason cannot substantiate speculation that ventures beyond the realm of experience (see, for example, A 244/B 303). However, this distinction opens up the possibility of believing in things the existence of which cannot be theoretically or empirically proven. After all, if transcendental idealism is true, experience does not teach us, for example, that freedom does not exist. Experience teaches us merely that freedom does not exist in the realm of experience.[32] We even know our own selves only a posteriori, under temporal conditions (IV 451.21–3). There is no privileged access to the self, as it is in itself. It is true, we cannot know with theoretical certainty that the practical standpoint that we feel we must take up in deliberation is not an illusion. But even if experience of human action displays the usual empirical regularities, moral philosophy can proceed to develop the hazy idea of an 'intellectual world' in which another self is active independently of empirical conditions, as long as it does not claim to possess cognition of corresponding objects.[33]

¶ IV 451.37 Common understanding is so enamoured with the world of sense that it tends to assimilate even the crude notion of an intellectual world to the conditions of sensibility. We are reminded of artistic representations of the soul as a little winged creature, or a whiff of breath, as it escapes a lifeless body.[34]

¶ IV 452.7 Drawing on the first *Critique*'s distinction between the understanding and reason, Kant now reveals why we must conceive of ourselves as members of an intelligible world: it is because we are rational beings. Rational activity cannot be merely passive. Reason, even more so than the understanding, goes *beyond* the world of sense by virtue of its capacity to develop ideas (see A 312/B 368 ff.). The reality of ideas that are entirely divorced from all empirical conditions is highly problematic, but this does not affect Kant's present argument.

32 That is the limited sense in which the antithesis side of the third antinomy is revealed to be correct (A 445/B 473); of course, its dogmatic proponent lacks the resources to distinguish between the realm of experience and things as they are in themselves, and thus concludes that if freedom cannot be encountered in experience freedom does not exist. See my 'Warum scheint transzendentale Freiheit absurd?', *Kant-Studien* 91 (2000), 8–16.

33 Kant first developed the notion of an intelligible world in his pre-critical inaugural dissertation on the form and principles of the sensible and the intelligible world (*De mundi sensibilis etc.*) of 1770, II 385–420.

34 At the other end of the intellectual spectrum, Kant accuses philosophers like Plato of making the reverse kind of mistake by abandoning experience altogether (IV 462.22–9). On the difference between the two worlds and its laws cf. IV 459.3–31.

For now, the only thing that matters is the fact that human beings are aware of a capacity within themselves to *think* beyond the boundaries of experience. When we conceive of these ideas we make what might be called a 'leap of reason'. Reason as pure spontaneity thus reaches far beyond the mere comparative spontaneity of the understanding, which is tied to empirical conditions (such as two events given in experience that are judged to be cause and effect). We should, however, note that our membership of a world of understanding could be confined to theoretical reason (see the scenario at IV 395.15–27). Neither the freedom to judge nor the freedom to think beyond the boundaries of the understanding implies that we possess a will that is free.[35]

¶ IV 452.23 Kant completes the argument for our membership of a supersensible world: the fact that in generating ideas as concepts of pure theoretical reason we dissociate ourselves, by virtue of our being 'intelligences',[36] from the world of sensibility can count as evidence. We can view the exercise of our practical faculties from two vantage points, and 'hence' (*folglich*, IV 452.27) our actions, which from one perspective display the regularities of nature – with reference to our will: heteronomy – and from the other are subject to the laws of autonomy.

¶ IV 452.31 When we think of ourselves as free we conceive of ourselves as part of an intellectual world and subject to its laws.[37] It is this assumption that all further inferences rest on. The judgement that we belong to a world of understanding is a synthetic proposition a priori, not simply an assumption that analytically follows from acknowledging the validity of moral laws.

The last three paragraphs of this subsection (IV 452.7–453.2) thus echo the previous discussion of rational spontaneity (IV 448.11–22), but the argument has moved on in two important respects. First, the notion of an intellectual world now enables us if not to understand, at least to conceive of a self that is radically independent of the conditions

35 See D. Henrich, 'The Deduction of the Moral Law: The Reasons for the Obscurity of the Final Section of Kant's *Groundwork of the Metaphysics of Morals*', in *Kant's Groundwork of the Metaphysics of Morals*, ed. P. Guyer (Rowman & Littlefield, 1998), pp. 314 and 319–20.

36 'Intelligence' (IV 452.23–4) in this sense is used synonymously with 'rational being' (see V 125.17). The 'lower faculties' (*untere Kräfte*, IV 452.24) are those belonging to the world of sensibility.

37 In fact, it ideally governs the actions of rational beings in just the same way in which natural causality determines appearances (see IV 446.7–12) – but of course we are not exclusively rational.

of the world of sensibility. This perspective was previously unavailable. Secondly, the context of the earlier passage was exclusively practical. The attribution of spontaneity was founded on the universal and necessary nature of the commands of morality, which apply to all rational beings as such; whereas the ideas of reason referred to at IV 452.18 to confirm our special status are concepts of pure theoretical reason. However, Kant has not yet shown that pure reason can be practical, i.e. how it can determine the will by itself and completely independently of inclination. We have discovered a standpoint from which we can escape the circle, and Kant has demonstrated that this standpoint is not as odd or unusual as one might at first think; but there is as yet no explanation of the synthetic nature of the categorical imperative. Nor can we account for the fact that we take an interest in morality.[38]

¶ IV 453.3 Kant now explains why circularity would have been fatal for the project of Section III. He is trying to corroborate the claims of morality. The first subsection contains sufficient proof that a will's being subject to moral laws – in the shape of the principle of autonomy – follows or 'flows' from the assumption of absolute freedom of the will (IV 446.14). If we can show that this presupposition is justified we have obtained an important intermediate result. We infer morality, as autonomy, from freedom.[39] If a circle was to be found *in that*, there would be no independent grounds for the presupposition of freedom and the project of Section III would fail. We would be begging the question. Kant would be in the position of a person trying to show that God exists on the grounds that the Bible contains testimony of his existence; and who, when asked to justify the authority of the Bible, says that everything in it is true because, after all, it is the word of God – both the supposed proof of God's existence and the problematic argument for the validity of morality are genuinely circular (or indeed the philosopher who relies upon the principle that clarity and distinctness convey truth to argue

38 That is why it would be a mistake to suppose that Kant is trying to give a deduction of human morality in purely theoretical terms. Mere membership of a world of intelligences cannot be equated with human autonomy. There is, at least theoretically, the possibility of rational beings endowed with theoretical reason only. Kant is trying to illustrate or confirm the idea of our membership in a world of intelligences by means of an example borrowed from theoretical philosophy, but the fact that we consider ourselves members of such a world when we act is still urged on us by our recognition of the validity of moral laws.

39 The final step of the argument would have to consist in the half of the circle introduced at IV 446–7, which is the inference from freedom to autonomy and, with recourse to Section II, from autonomy to morality, as summarised at IV 453.4–5.

for the existence of God, and then claims that God is the guarantor of all his clear and distinct ideas).[40]

We cannot, when we think or act, accept that our activity is merely the result of natural causation; to the contrary, we must assume that causal regularities can be appearances of rational influence. Conceiving of ourselves as rational thinkers and actors releases us from the world of sense. We 'transfer' or 'transport'[41] ourselves into a realm governed by the laws of reason. In the Methodology of the first *Critique* Kant makes a similar point. There, surprisingly and controversially, he tries to divorce transcendental and what he calls 'practical freedom'; in this case, freedom of action. The latter, he argues, 'can be proven through experience'. Kant then introduces the notion of imperatives, defined as 'objective *laws of freedom*', which tell us '*what ought to happen*, even though it never does happen' and distinguishes them from '*laws of nature*, which deal only with that *which does happen*'. Whether our actions are ultimately determined by natural forces is something that 'in the practical sphere does not concern us, since in the first instance we ask of reason only a *precept* for conduct' (A 802–3/B 830–1).

Kant's point is that the prescriptive question of *what I ought to do* is different in kind from the descriptive question of what *I am doing*. Natural determinism is a threat to moral responsibility, but it is irrelevant from the first-person point of view of deliberation. He makes essentially the same point in his review of Johann Heinrich Schulz's *Attempt at the Teaching of a Doctrine of Morals*, published in 1783. Transcendental and practical freedom are distinct; we bracket the former when it comes to action; everyone, even a fatalist, must act as if he were free (VIII 13.20–33). In the *Groundwork*, this phenomenon is taken to be an indication that not all of human nature is subject to the laws of cause and effect, which – on the background of Kant's agnosticism about freedom in the theoretical sphere – provides a new vantage point from which we judge ourselves to be free. Transcendental freedom has its role to

40 The two circles differ in that Kant's argument seemingly contains two analytic proofs (whereas at least the argument for God's existence rests on the empirical inspection of a book called the 'Bible'). If the danger of the circle was real, we would not be in a position to account for the categorical imperative as a synthetic practical principle. We would still be caught up in conceptual analysis. But now the notion of the 'other' point of view that we take up as rational beings has shown a way out of the analytic game. Kant can now proceed to show that of the two points of view our membership of an intellectual world is more important.

41 *Versetzen*, IV 453.12. The image is taken up again at IV 454.32, IV 455.2 and IV 457.9; see also second *Critique*, V 42.19 and V 43.30.

play in moral philosophy; but it is simply irrelevant when it comes to deliberation.

On the present interpretation, the argument of Section III so far is compatible with the footnote at V 4 of the *Critique of Practical Reason*.[42] Kant pre-empts the charge of an alleged 'inconsistency'. Someone might ask how it is possible that freedom is the condition of the moral law and the moral law the condition of one's consciousness of freedom. Kant's reply: freedom is the *ratio essendi*, the sufficient ground of being, of the moral law; whereas moral conviction is the reason why we are aware of our freedom, its *ratio cognoscendi*. The footnote can be read as a clarificatory summary of Kant's escape from the circle. When we consider ourselves free, we do not analytically infer that we are free on the grounds that we recognise the force of morality. Rather, moral consciousness propels us out of the world of sensibility and forces us to take up the point of view of rational agency. (Note also that our consciousness of freedom has – tentatively – been explained, which is reassuring; but there is still no strict proof that human beings possess a free will.)

4 *The 'deduction': how is a categorical imperative possible?*

¶ IV 453.17 Kant returns to the main purpose of Section III: the 'deduction' of the categorical imperative. The preparatory distinction between a world of sense and a world of understanding, as drawn in the course of resolving the circle, is now put to good use. In the *Critique of Practical Reason*, 'deducing' an a priori concept – such as the concept of natural causation – is said to be required because of the necessity of the combination that it entails; and amounts to a demonstration of 'its possibility from pure understanding without empirical sources' (V 53.30–2). Kant now attempts to do this for the categorical imperative.

We have established that a rational being in general (*das vernünftige Wesen*, IV 453.17) – to be precise, a finite rational being like us – must consider himself a member of both the world of sense and the world of understanding. He *counts* himself a member of the intelligible world and as possessing a free will; but he is *conscious* of being a member of the sensible world, in which the effects of free volition surface as

42 In fact, this seems to be Kant's position throughout the critical period. However, in the *Critique of Practical Reason* the *ratio cognoscendi* is narrowly moral, whereas in the *Critique of Pure Reason*, A 547/B 575 (cf. A 802–3/B 830–1) it is practical in a broader sense of the word, i.e. it includes action on hypothetical imperatives.

'mere appearances' (IV 453.21). There is no trace of the supernatural provenance of human action in the phenomenal world. Rather, actions appear to be determined by other appearances – desires and inclinations – in accordance with the causal laws of nature. The causality of freedom lies beyond the realm of knowledge. Even so, the assumption of our dual membership remains.

If – like God or non-rational animals – we were members of just one of these two worlds, we would be subject to only one type of legislation and there could be no room for conflict. The question whether the one law or the other is more authoritative, and if so why, does not arise. We would act for the sake of the moral law if we were pure intelligences; for the sake of our natural desires – and ultimately our own happiness – if we were bodily creatures only (IV 453.30–1). The problem of the authority of moral laws – the problem addressed in this paragraph – arises because we are members of *both* worlds. We would like to know why the legislation of reason is superior to the legislation of nature when the two come into conflict with each other.[43] You are, as it were, members of two different clubs, the Club of Sensibility and the Club of Understanding, and subject to the regulations of both – say to wear their respective tie or scarf when you attend their events. This is unproblematic for those of your friends who are members of one club only. But for you, an invitation to a *joint* event creates a clash of sartorial rules: which tie or scarf should you wear? You are conscious of your membership of both clubs. You may be certain that you ought to be loyal to the Club of Understanding, but you have not yet been presented with an account of the greater authority of the regulations of either society. Moreover, the tie of the Club of Sensibility is *so* much more elegant.

Kant proceeds to offer an argument for the primacy of the intelligible world that is sketchy at best. It is considered to be authoritative within the human will because, if we possess transcendental freedom, it can actively *determine* the laws of the world of sense. The world of understanding, Kant argues, 'contains the *ground* of the sensible world, and

43 Kant makes it quite clear that he is concerned with the conflict of two distinct reasons for action: morality and happiness. He is not arguing for the authority of imperatives as rational commands in general, he is giving an account of the authority of pure practical reason. The contribution of empirically directed, in fact theoretical, reason towards action on hypothetical imperatives is minimal. The only imperative that is relevant at this point in the argument of Section III is the imperative of duty (see IV 454.4–5). For a particularly clear statement of the problem see Hill, 'Editorial Material', p. 103; see also his 'Kant's Argument for the Rationality of Moral Conduct', in *Dignity and Practical Reason* (Cornell University Press, 1992).

therefore also of its laws' (IV 453.31–3). In short: a moral agent is the origin of the laws that describe his action at the level of appearances. By contrast, the world of sense cannot necessitate the intellectual world. We have already seen that we must consider ourselves free from its influences in rational activity. The two worlds do not possess equal status. It is this fundamental asymmetry that enables Kant to conclude that in cases of conflict between the recommendations of inclination and the commands of pure practical reason we always have to side with the latter. The will wholly belongs to the superior world of understanding (IV 453.33–4), and it is the laws of this world that gives the will its proper law. That is why, for the joint garden party, you must wear your drab *Verstandeswelt* tie.[44]

¶ IV 454.6 However, the central task of *Groundwork* III still lies ahead. The deduction is finally accomplished in the present paragraph.[45] Its first sentence emphatically proclaims that 'categorical imperatives are possible as follows'; and the next paragraph begins with the assertion that 'this deduction' is confirmed by the practical use of common human reason (IV 454.20–1). Kant has established the origin, relevance and authority of moral laws, but we still need to know *how they connect to the human will*.

The deduction returns to the idea that synthetic propositions, in the theoretical as in the practical realm, require 'a third something', a representation fit to combine two essentially separate concepts (one concept does not contain or analytically follow from the other). Kant also states the synthetic proposition he seeks to corroborate: an absolutely good will is a will whose maxim can always contain itself within itself, regarded as a universal law; in short, an absolutely good will is a will that acts in accordance with the principle of autonomy (IV 447.10–19). If we presuppose, as we must, that human beings ought to aspire to absolute goodness – even if moral action is not done for the sake of some perceived good – this principle can be expressed as an imperative:

44 Kant's definition of the will at IV 412.26–8 contained a first hint that human beings are capable of shaping natural regularities (as did the law-of-nature variant of the categorical imperative). For a recent argument to this effect see E. Watkins, *Kant and the Metaphysics of Causality* (Cambridge University Press, 2005), esp. pp. 257–65.

45 *Pace* D. Schönecker, who takes the deduction to be Kant's argument for the authority of the law of autonomy and therefore locates it in the previous paragraph (*Kant: Grundlegung III*, p. 365). But the central question, phrased in terms of the possibility of the categorical imperative as a synthetic a priori principle throughout, is answered only in the present paragraph; and it is more plausible that the opening line of the next paragraph ('this deduction', IV 454.20) refers to the *immediately* preceding paragraph.

always act in accordance with the principle of autonomy. This much is clear.

The minutiae of the deduction are rather less obvious, but the wording of the present passage recommends the following reconstruction. To establish the thesis that my finite will – and by extension the will of any other rational being like me – ought always to act in accordance with the principle of autonomy, I need to combine the following two separate concepts: (i) *my human will*, which is affected by sensuous desires (IV 454.12), and (ii) *the laws of autonomy* (IV 454.8–9). The connecting concept is (iii) the idea of the very same will, i.e. my will, *as pure and by itself practical*, residing in the world of understanding (IV 454.13–14; its provenance was established at IV 453.33–4).[46] As a result, the laws of the intellectual world – autonomy – must be the condition of all actions of 'the former' (*des ersteren* IV 454.14–15), i.e. of my will. If my will were at home in the intelligible world only, all actions would be in accordance with the laws of autonomy. As I am also a member of the world of sensibility, I experience the moral law as an imperative.[47]

The deduction is problematic in at least the following three ways. First, Kant thinks that the two concepts to be conjoined in a synthetic judgement must both be combined in the third element responsible for the synthesis (see IV 447.16). In the present case, (iii) the idea of a pure will located in an intelligible realm arguably contains (ii) the laws of autonomy; but it is much less obvious that it also contains (i) the idea of itself as a finite will. Secondly, in the *Critique of Pure Reason* the minimal condition of the objective validity of synthetic judgements is time. A synthetic judgement cannot be objectively valid if the 'third something' is not given in intuition, such as the 'concept of reason' (idea) of a pure will. In light of this, the deduction appears to be less than successful.[48] Thirdly, some interest in moral behaviour is needed

46 *Die Idee ebendesselben* [*sc.* of my very own will, IV 454.12], *aber zur Verstandeswelt gehörigen reinen, für sich selbst praktischen Willens*. A pure will or the intelligible world as such cannot make this connection. The 'third something' needs to be a *representation*, i.e. the idea of such a will. See also the 'combination' effected by the idea of personality in the *Critique of Practical Reason*, V 86.36.

47 The phrase 'cognition of a nature' (*Erkenntnis einer Natur*, IV 454.19) is odd. 'A' or 'one' nature as opposed to another? If so, is he saying that there are two kinds of nature, both of which are apprehended by means of synthetic propositions a priori? Or is he using 'nature' as a synonym of 'creature' again, in which case the nature – a being – would be the cogniser by means of synthetic propositions a priori? On the former reading, we can acquire knowledge of the world of understanding, even though we cannot know that its laws can determine the human will in action.

48 By way of illustration, Kant at IV 454.15–19 rather vaguely (*ungefähr so*) refers to the transcendental deduction of the categories in the *Critique of Pure Reason* (A 84/B 116 ff.).

to make the connection in practice, which is duly posited: reverence for the moral law. However, the possibility of explaining this interest lies beyond the confines of practical philosophy (see IV 459.32–460.7 below).

A fourth – even more fundamental – difficulty results from the fact that two worlds or vantage points do not suffice to provide us with a credible explanation of the possibility of a categorical imperative. If Kant assumes that the will wholly belongs to the world of understanding and consequently is free to act according to its laws, it seems quite impossible to understand the reasons why human beings ever fail to live up to these standards. The mere fact that the human will is *affected* by sensuous inclinations cannot be the (whole) reason for occasional moral failure, for the will would not otherwise be (negatively) free. Yet if it is (positively) free to do the right thing it is – as Kant at times admits – unintelligible why it does not always realise its freedom. The executive will (*Willkür*) would have to occupy a third standpoint mediating between the two others; and this standpoint is as uncertain as that of empiricism (IV 425.33), perhaps even more so.[49]

¶ IV 454.20 Common practical reason *confirms* the results of our deduction: every human being, even the most perverted villain (*der ärgste Bösewicht*, IV 454.21–2), naturally considers himself part of the intellectual world and identifies with his higher self, not his natural desires. Kant's optimism is astonishing. When presented with examples[50] of decent conduct, Kant avers, the villain develops a fervent wish to become a good person, but deep-rooted sensuous inclination makes it extraordinarily painful for him to bring this change about.[51] Yet he is convinced that it can be done. In deliberation, even he considers himself free and subject only to his own rational laws. The villain's moral consciousness is said to be *proof* of his possessing a pure will with which he can 'transfer' himself into that other realm (IV 454.29–30). The similarities with the *Critique*'s 'fact of reason' are obvious. In the *Groundwork*

49 See Appendix E for a discussion of the related problem of moral failure.

50 See second *Critique*, V 92.14. This is a legitimate use of examples in moral practice. On the limits of an ethic of role models see IV 408.28–33 above.

51 The villain 'cannot easily' (*kann . . . nicht wohl*, IV 454.27–8) bring this about; not to be confused with 'probably not' (*wohl nicht*). *Wohl*, like its English equivalent 'well', is the adverb corresponding to *gut*. In Kant's eighteenth-century German *wohl* has connotations of pleasantness and facility. Kant does not say that the worst villain cannot become a decent person, as his English translators would have it.

already there is no getting away from the 'commonest use of practical reason'.[52]

Furthermore, it seems that common moral consciousness is needed to identify acting according to the laws of the intelligible world as morally *good*. Value terms are conspicuously absent from the deduction so far. They are re-introduced only in the present paragraph: at IV 454.24 we learn that the villain approves of those who resolutely act on 'good maxims'; and at IV 455.4 he is said to be conscious of his own 'good will'. In fact, the *Groundwork* presents itself expressly as an inquiry into moral value, even if eventually it turns out that it must be defined in purely formal terms. Kant starts with the notion of moral goodness; he returns to it in his overview of the different formulations of the categorical imperative towards the end of Section II (IV 437.5–440.13); and the principle to be deduced in Section III is explicitly formulated in terms of *good* volition (IV 447.11). Kant has made good his promise to return to his starting point. The deduction of autonomy extends just as far as it is supported by the 'practical use of common human reason'. This must be at least part of the reason why Kant can reject the possibility of a deduction of the categorical imperative in the *Critique of Practical Reason*.[53]

5 *The extreme boundary of all practical philosophy*

a. The problem of reconciling natural necessity and free will does not yet mark the extreme boundary of practical philosophy

¶ IV 455.11 Kant turns his attention to the concept of freedom. The reason is that it points to the extreme boundary of moral philosophy that he now intends to demarcate. This is made explicit only at IV 461.7–14. We can state the sufficient condition of autonomy: freedom, and the presupposition this entails – our membership of an intellectual

52 Ordinary moral consciousness is also invoked in the later work; see e.g. V 91.20, V 32.2–7 and V 43.36–7.

53 For Kant's scepticism, in comparison with the deduction of the categories, see *Critique of Practical Reason*, V 46.20–4. See also V 87.31–5, V 93.30–94.7 and V 105.10–13. The present reading suggests that Kant has not so much changed his substantive views about human autonomy and its intelligible ground but rather his opinion as to whether the actual deduction at IV 454.6–19 is a success. A possible explanation might be that while he was revising the *Critique of Pure Reason* for its second edition he was reminded of the minimal condition for objective validity that in the case of synthetic a priori judgements the third something must – minimally – be given in time. The chapter on synthetic judgements is reprinted without major changes (B 193–7). In the second *Critique*, objective validity is indeed declared to be the stumbling block; see V 46.20–4.

world. But the deduction cannot proceed any further than that (see IV 463.21–33).

Our experience extends only to actions that lie in the past; and acts, like all events, happen in accordance with the laws of nature. Yet the *necessity* with which the same cause produces the same effect, though confirmed by experience, is not empirical in origin. In the transcendental deduction of the first *Critique*, Kant claims to have shown that our use of synthetic judgements a priori in applying categories like causality is justified. The categories are pure concepts of the understanding that shape human experience.

By contrast, the principles that determine what we do – our maxims – are not empirically accessible. If they are freely chosen they cannot, by definition, be part of the web of cause and effect even if the acts they produce are. We can hardly imagine the experience of the fact that an action could have happened, or ought to have happened, when in fact it did *not* happen. Freedom is a concept not of the understanding but of reason, an 'idea', and as such distinct from the realm of experience. We have been presented with an account of why it is the case that we consider ourselves free; but this leaves the problem of how – their different levels notwithstanding – causality as a concept of the understanding can be compatible with freedom as an idea of reason. In what follows, Kant summarises the paradox of the *Critique*'s third antinomy and its resolution.[54]

¶ IV 455.28 The dialectic of reason with regard to the principle of causality arises as follows. On the quest for a sufficient explanation of natural events, reason pursues two strategies that, ultimately, are equally unsuccessful: the search for a free first cause to explain all subsequent events and the infinite regress within a series of homogeneous natural causes. There appears to be a contradiction because one and the same event must be *both* part of a natural process *and* the effect of an act of freedom, which entails independence of natural laws (A 444/B 472). However, the antinomy is declared to be a problem not of practical but of theoretical reason, which is closely tied to experience and therefore favours the second strategy. Yet freedom is supported by an equally rational practical (moral) interest.

54 The other *locus classicus* is the 'critical elucidation' of the Analytic of the *Critique of Practical Reason*, V 89–106.

Compared to the first *Critique*, Kant's account in Section III of the *Groundwork* is playing down the theoretical interest in freedom of the will. It almost seems as if there was a dialectical opposition between theoretical and practical reason, rather than a conflict within theoretical reason. In the *Critique of Pure Reason*, practical reason is merely said to support the assumption of freedom (A 466/B 494).

¶ IV 456.7 Natural necessity stands on firmer ground because it is validated by experience; and it was previously established by a transcendental deduction. If there was an irresolvable conflict, freedom would have to give way. That is why the 'seeming contradiction' of the antinomy must at least be resolved (cf. B xxix).

¶ IV 456.12 Theoretical philosophy paves the way for morality by accounting for the transcendental illusion that freedom and natural causality are incompatible. They are assigned different points of reference. As appearances, human beings are subject to natural laws; as members of the world of understanding they are free; and the latter sphere is prior because it determines the regularities of the former. Kant further argues that philosophers do not have a choice as to whether to resolve the antinomy or not. As long as there is no satisfactory answer the opponents of freedom have won. The problem of free will would be a good that has no apparent owner, a *bonum vacans* (IV 456.31).[55] The 'fatalist' (determinist) side could claim it because from the point of view of natural science everything counts in favour of the principle of causality, and nothing in favour of freedom.[56]

¶ IV 456.34 As the reconciliation of freedom and causality is part of *theoretical* philosophy, it cannot mark the outer limit of *practical* philosophy. (It would be more appropriate to say that it marks the beginning.) The problem of free will and determinism does not concern the substantive question of what we ought to do, even though a certain answer must be presupposed for moral philosophy to be possible as a normative and action-guiding discipline.

55 Kant returns to the legal language of a 'deduction', the purpose of which is to *substantiate* a claim (see also IV 457.4). See Henrich, 'The Deduction of the Moral Law', p. 323.

56 See the summary in the second-edition preface of the *Critique of Pure Reason*, B xxvii. We may well have our doubts whether determinists will find transcendental idealism persuasive. Of course, Kant assumes that even his empiricist opponents have an interest in the reality of freedom, not as speculative philosophers but as moral agents, and will therefore be relieved to hear that there is a way out of the conflict that satisfies both sides.

b. We are conscious of our free will but cannot cognise or explain it

¶ IV 457.4 The 'legal claim' or 'entitlement' (*Rechtsanspruch*) to freedom is grounded in our being conscious of the independence of reason. Returning to moral philosophy proper, Kant summarises the results of the deduction (IV 453–5) before he focuses his attention on its limits.

¶ IV 457.25 It is because we are conscious of ourselves as intelligences that we conceive of ourselves as responsible for our actions. Kant says that man (*der Mensch*) has a will that lets nothing stick to it ([*läßt*] *nichts auf seine Rechnung kommen*, IV 457.26) that merely belongs to his inclinations. We are persuaded that we always *could* have done the right thing even if we followed inclination when we should have acted from duty. Yet, as Kant argues in the *Critique of Practical Reason*, we are conscious of our freedom – with its very own laws – when we face the categorical commands of duty – which is also implied at IV 410.28–9 above. As the law of autonomy is our very own law, whereas the law of nature is not, our 'real' or 'proper' self (*das eigentliche Selbst*, IV 457.34; cf. IV 458.2) is declared to be that which belongs to the intelligible world (the 'intelligence', i.e. the rational element in isolation). The human being we are familiar with – in fact, the manner in which we *experience ourselves* – is merely the self as it appears.

¶ IV 458.6 We are entitled to our thoughts as long as they conform to the categories, even if they are removed from factual experience; all the more so if our thoughts are driven by (practical) reason. However, we must give up any claim to cognition of or direct exposition to an intellectual world because it lies beyond what we can experience (see Kant's warnings at *Critique of Pure Reason* B xxvi fn. and B 166 fn.). More needs to be said about Kant's thesis that the boundaries of practical reason would be transgressed if it were to take an object or motive out of the world of understanding. He is probably alluding to the thought explained at IV 449.13–23 that no interest rooted in the world of understanding compels us to take up the moral point of view; in fact, even the interest we must necessarily take in moral action remains inexplicable.

¶ IV 458.36 Kant now turns to practical interest. This paragraph is where, strictly speaking, the 'extreme boundary' begins. The question of the possibility of freedom is more specifically equated with that of

how the moral law can directly determine the will in the second *Critique*, V 72.21–4.

¶ IV 459.3 Kant identifies the concept of explanation – somewhat artificially, but in accordance with the findings of the *Critique of Pure Reason* – with that which can be traced according to the laws of experience. As free action is independent of the laws of nature, it is trivial that it cannot in this sense be explained. If free actions could be explained, not as mere events of the physical world but as actions, they would not be free. That which makes them free human actions can never surface in this way. The attempt thus to explain free actions involves a category mistake similar to the tendency of making everything intellectual 'sensual' mentioned above (IV 452.4–6).[57] Practical reason retains the task of defending its claims to the best of its ability, as in this last section of the *Groundwork*.

c. The inexplicability of the interest we take in morality is the outermost boundary of moral philosophy

¶ IV 459.32 Today's readers face the difficulty that the meaning of 'moral feeling' (*moralisches Gefühl*) is not entirely obvious. It designates both a feeling or sentiment in the modern sense and the moral sense, i.e. a capacity like the other senses of hearing (*Gehör*), taste (*Geschmack*), smell (*Geruch*) and sight (*Gesicht*).[58] Kant rejects the thought that such a capacity and the resulting sentiment could be the foundation of moral judgement. If we possess a feeling that is in harmony with morality, it must itself rest on the unconditional judgement of reason. The problem of a felt interest, which we cannot explain, is thus identical with the question of whether our will is free because action independent of all inclination or sensuous interest – i.e. free action – is possible only if reason with its judgement generates its own interest. This interest was already equated with reverence for the law in Section I (IV 401 fn.).

Fn. IV 459–60 Kant defines the term 'interest' and emphasises that in the case of moral volition one has a pure interest in *willing* the action, not directly in the realisation of that which the action aims at.

57 On the analogies of experience see *Critique of Pure Reason*, A 176/B 218 ff. On the category mistake involved in trying to explain free action see also Kant's preparatory notes on *Theory and Practice*, XXIII 141–2.

58 'Feeling' can, of course, have this meaning – cf. Macbeth addressing the dagger: 'Art thou not, fatal vision, sensible to feeling as to sight?' (II.1.36–7).

Essentially the same point is made in a previous footnote in Section II (IV 413–14 fn.). Recall the 'second proposition': an action is not good because of the purpose pursued (IV 399.35–7). If I intend to help someone *from the motive of duty*, reverence for the moral law, I have an interest in an act of volition and the action proceeds on this basis. By contrast, if I intend to help someone from sympathy, I merely will that the person be helped, which could equally be effected by chance or by a mere natural cause. The question left open even in Section III can be reformulated as follows: how is it possible that I *will* that I take an interest in something, even if I have no interest in the *thing* that I will?

¶ IV 460.8 The will requires a mediating interest or motive even in completely free action determined by the moral law. The way in which *sensuous* stimulation provokes a 'pathological' interest can be explained physiologically. By contrast, it is impossible to cognise how the thought of *moral* necessity, the universality of a law of reason, or the idea of a world of understanding should create, in moral judgement, an interest to do the moral thing *ex nihilo*. The cause of this interest is not an event within the world of experience. We must accept it as given.[59]

¶ IV 461.7 Freedom is the condition of morality. We must consider ourselves free beings, because we are conscious of ourselves as members of a world of understanding. The deduction is limited, but it suffices to meet the challenge of empiricism and thus to justify the moral theory explicated in Sections I and II from a *practical* point of view. However, freedom not just in judgement but also in action is possible only if we take an interest in moral commands; and we do take an interest. Yet it is given and cannot be explained because it does not arise naturally. In moral action we pursue ends that do not rest on natural incentives. The interest is merely *practically* necessary.

¶ IV 461.36 Kant intimates that the problem of explaining a moral interest is connected with the formal character of the moral law. Explanation rests on matter, something given in intuition. Over and above

59 Kant again argues that everything sensuous must be subordinated to what is rational; see IV 453.31–454.5 above. It is only thus that Kant can justify his thesis that the moral imperative commands with absolute necessity and does not just supply reasons that can be weighed against other competing reasons. Compared to duty, all other reasons at once lose all their value; they can only be acted on within the limits of moral permissibility.

our moral convictions we are not justified in saying anything about a 'world of intelligences'.

¶ IV 462.22 This summary touches on two famous themes of the *Critique of Pure Reason*. First, Kant warns us not to leave the world of experience like Plato on the 'wings of the forms' (or 'ideas') hoping to make better progress in metaphysics in the world of understanding (see A 5/B 9 and *Critique of Practical Reason*, V 141.9). The danger of losing one's way in groundless speculation can be averted in the field of practical philosophy because of the given but inexplicable interest in morality. Secondly, Kant hints that knowledge in matters of freedom must be removed (*aufheben*) in order to 'make room for belief' (B xxx), as he does in the second edition of the Preface. This belief is not the bookish belief of revealed religion but rather a practical belief of reason.

6 Conclusion: Comprehending that we cannot comprehend morality

¶ IV 463.4 It is a distinguishing feature of our faculty of reason that it always strives to pursue its projects to their final and necessary conditions. Theoretical reason thus produces the idea of a first free cause as the ultimate explanation of the world.[60] Similarly, practical reason arrives at the notion of a categorical, i.e. unconditional, imperative as the ultimate principle to which all human action must conform.

In both cases the results remain somewhat inconclusive. The notion of a first cause extends beyond experience and, at least from the point of view of theoretical reason, is no more than a mere idea. As the first principle of practical philosophy, the categorical imperative is unconditional in the sense that it applies to all rational beings as such, independently of their regular desires. Any such condition would make the imperative hypothetical. However, in Section III the metaphysical – not practical – condition of a categorical imperative has been traced back to the consciousness of freedom, which is revealed to us in practical deliberation. We are unable to explain how our interest in morality – and hence our consciousness of our radical independence – comes about. We would, of course, like to investigate the reasons or grounds for this assumption too. But as necessity can only be explained by pointing to

60 On the thesis side of the third antinomy, see *Critique of Pure Reason*, A 444/B 472.

an equally necessary condition of the fact to be explained, we have now reached the end of the chain of conditions. The regress – a regress not about the conditions of value, but about the conditions of rational action – must come to a stop somewhere.

Ultimately, the 'proof' of Section III therefore amounts to a reassuring inference to the best explanation of the phenomenon of morality against the very strong claims of empiricism. It is an account that human reason willingly accepts. Autonomy is traced to freedom; freedom presupposes membership in a world of intelligences, an assumption we make but cannot corroborate. Freedom cannot be explained. We can comprehend not morality but at least the limits of moral philosophy.

BIBLIOGRAPHY

Ameriks, Karl. 'Kant's Deduction of Freedom and Morality', in *Interpreting Kant's Critiques* (Clarendon Press, 2003), 161–92 (first published in 1981)

'Kant's *Groundwork* III Argument Reconsidered', in *Interpreting Kant's Critiques*, 226–48 (first published in 2001)

Brandt, Reinhard. 'Der Zirkel im dritten Abschnitt von Kants Grundlegung zur Metaphysik der Sitten', in *Kant. Analysen – Probleme – Kritik*, ed. H. Oberer and G. Seel (Königshausen & Neumann, 1988), 169–91

Henrich, Dieter. 'The Deduction of the Moral Law: The Reasons for the Obscurity of the Final Section of Kant's *Groundwork of the Metaphysics of Morals*', in *Kant's Groundwork of the Metaphysics of Morals*, ed. P. Guyer (Rowman & Littlefield, 1998), 303–41 (first published in German in 1975)

Hill, Thomas E. 'Kant's Argument for the Rationality of Moral Conduct', in *Dignity and Practical Reason in Kant's Moral Theory* (Cornell University Press, 1992), 97–122

Korsgaard, Christine. 'Morality as Freedom', in *Creating the Kingdom of Ends* (Cambridge University Press, 1996), 159–87

McCarthy, Michael. 'Kant's Rejection of the Argument of *Groundwork* III', *Kant-Studien* 73 (1982), 169–90

'The Objection of Circularity in *Groundwork* III', *Kant-Studien* 76 (1985), 28–42

O'Neill, Onora. 'Reason and Autonomy in *Grundlegung* III', in *Constructions of Reason* (Cambridge University Press, 1989), 51–65 (first published in 1989)

Timmermann, Jens. 'Warum scheint transzendentale Freiheit absurd?', *Kant-Studien* 91 (2000), 8–16

Appendix A: Schiller's scruples of conscience

Whether Friedrich Schiller's distichs represent a serious attack on Kant's ethical theory or a parody of its critics,[1] it is a solemn duty to discuss his verses in a commentary on the *Groundwork*:

> *Scruples of Conscience*
>
> Gladly I serve my friends, but alas I do it with inclination
> And thus I am frequently nagged by my lack of virtue.
>
> *Decision*
>
> There is no other advice, thou must seek to despise them,
> And do with disgust what thy duty commands.[2]

As a criticism of Kant's moral psychology Schiller's advice would be thoroughly misguided. First, a terminological point. No-one has ever reason to regret doing anything 'with inclination' if this expression is taken to indicate the mere presence of inclination, rather than its determining the course of action. The presence of concurrent inclination does not annihilate the moral worth of an action *if* it is done for the sake of duty. Schiller presumably wishes to say that, regrettably or not, he serves his friends 'from inclination'. Secondly, and relatedly, Kant does not believe that only actions done with disgust can be morally valuable. Even if we concede that overcoming subjective obstacles is particularly virtuous or admirable, we always have reason to prefer a friendlier, more harmonious inclinational state. Moral value need not be maximised. This is a direct consequence of Kant's doctrine of the (comprehensive) highest good: morally

1 Allen Wood plausibly defends the latter view (*Kant's Ethical Thought* (Cambridge University Press, 1991), pp. 28–9). If so, the character called 'Schiller' in our discussion of the distichs should not be confused with the German poet of the same name. On Kant and Schiller see also Paton, *The Categorical Imperative*, pp. 48–50, Dieter Henrich, 'Das Prinzip der kantischen Ethik', *Philosophische Rundschau* 2 (1954–5), 29–34, Allison, *Kant's Theory of Freedom*, pp. 180–4, Hill, 'Editorial Material', pp. 31–3, and Marcia Baron, 'Acting from Duty' in Allen Wood's translation of the *Groundwork*.

2 The German original reads as follows: 'Gewissensskrupel / Gerne dien ich den Freunden, doch tu ich es leider mit Neigung, / Und so wurmt es mir oft, daß ich nicht tugendhaft bin. // Decisum / Da ist kein anderer Rat, du mußt suchen, sie zu verachten, / Und mit Abscheu alsdann tun, wie die Pflicht dir gebeut.' Friedrich Schiller, *Werke* (Hanser, 1987), vol. I, pp. 299–300. The translation is my own.

good agents *deserve* to be happy. In the *Critique of Practical Reason*, he intimates that the taming of inclination might be an indirect effect of firm moral principles (V 117–18). Thirdly, the ideal of friendship, dubbed the 'hobby-horse of novelists' in the *Metaphysics of Morals* (VI 470.18), arguably lies beyond duty, especially if we go so far as to 'serve' our friends; which is amiable, no doubt, but not obligatory, and morally laudable only in so far as ethical restrictions are observed. If so, 'Schiller' might even be perfectly justified to serve his friends 'from inclination'. But he should not expect to receive any particular moral credit for it. Fourthly, even the worst caricature of Kant's theory of moral motivation does not warrant the conclusion that we should seek to 'despise' (*verachten*) our friends. We morally disapprove of those we despise, but for the most part our friends have done nothing to deserve this. Anyone so misguided as to wish to maximise virtuous behaviour could perhaps try not to despise but to *hate* his friends.

However, we should not let these famous verses obscure a very real difference between Kantian moral philosophy and broadly virtue-ethical approaches. In his *On Grace and Dignity* (1793), Schiller espouses the ideal of a harmonious 'beautiful soul' (*schöne Seele*), an agent for whom moral action has become second nature.[3] A beautiful soul is someone who performs moral actions effortlessly and gracefully because he or she has cultivated an 'inclination to duty'. If duty and inclination are in perfect agreement, the moral law ceases to be imperatival.

Kant must reject Schiller's ideal for two reasons. First, we can never be entirely certain that friendly inclination will in effect coincide with the demands of duty. It is impossible even in principle for finite beings like ourselves to achieve the ideal of a well-rounded moral character.[4] Secondly, Schiller's ideal of moral perfection is not just unrealistic. Nature, even if *per impossibile*

3 See Schiller, *Werke*, vol. V, pp. 433–88.

4 Chapter III of the 'Analytic' of the *Critique of Practical Reason* contains a thorough discussion of the misguided aspiration to holiness of character. Re-stating the distinction between actions that are done for the sake of the law (dubbed 'morality') and those that merely coincide with it ('legality'), Kant once again emphasises the radically different nature of moral and non-moral motives. He declares it to be 'of the greatest importance in all moral assessments to attend with the utmost exactness to the subjective principle of all maxims, so that all the morality of action is placed in their necessity *from duty* and from reverence for the law, not from love and affection for what the actions are to produce' (V 81.20–5). For 'created' beings like us, 'moral necessity is necessitation, i.e. obligation (*Verbindlichkeit*), and every action based on it is to be represented as duty, not as a kind of conduct which we already favour of our own accord, or could come to favour – as if we could ever reach a state in which without reverence for the law, which is connected with fear or at least apprehension of transgressing it, we of ourselves, like the Deity raised beyond all dependence, could come into possession of *holiness* of will by an accord of will with the pure moral law becoming, as it were, our nature, an accord never to be disturbed (in which case the law would finally cease to be a command for us, since we could never be tempted to be unfaithful to it)' (V 81–2). Moral legislation is directed at the attitude (*Gesinnung*) of the agent, not at mere acts. Kant calls beneficent action from love 'beautiful' but not morally good (V 82.18).

cultivated to perfection, prompts us to action that *conforms* with duty but is *not done for its sake*. As Kant puts it in the second *Critique*:

> [A]n inclination towards what conforms with duty (e.g. to beneficence) can indeed greatly facilitate the effectiveness of *moral* maxims, but cannot produce any. For in these everything must be directed to the determination of the law as determining ground if the action is to contain not merely *legality* but also *morality*. (V 118.9–14)

There can be no such thing as an inclination to duty, or a feeling attuned to morality (see V 75.28 and A 807/B 835). Inclination at best leads to action that coincides with duty. The two kinds of interest are different in kind. Even when reason and inclination recommend the same act, inclination inspires an interest in an outcome, not in the act. Kant was a dualist, Schiller a monist; no matter how their ethical theories are spelt out in detail, it is this fundamental disagreement that divides them.[5]

5 In *Religion within the Limits of Reason Alone*, Kant plays down the difference between Schiller's position in *On Grace and Dignity* and his own (VI 23–4 fn.). He unsurprisingly affirms, contra Schiller, that owing to the unrelenting necessity of its commands morality lacks 'grace' (*Anmut*), but – equally unsurprisingly, at least if the above analysis is correct – Kant expressly denies advocating an unhappy 'carthusian' frame of mind. After all, the moral law is a law we impose upon ourselves. It may be alien to inclination, in principle and occasionally in effect, but it is grounded in what we must regard as the part of our self that deserves priority (see e.g. IV 453–4).

Appendix B: The pervasiveness of morality

How extensive is morality? Must every deed be done 'from duty'? Should all actions possess moral worth? On closer inspection, these questions are not as straightforward as they first appear because the correct answer crucially depends on the identity criteria one applies to actions.[1] On the one hand, Kant makes it quite clear that no action can be entirely exempt from the moral law – see, for example, his dismissive discussion of morally indifferent actions at VI 23 fn. in his *Religion*. If so, all actions ought to possess moral worth in the weak sense that human beings must always pay due attention to the restrictions imposed by the moral law. It is never acceptable to act on a maxim that is altogether devoid of moral content. This does not, on the other hand, entail that every single action must have moral worth in a stronger and perhaps more obvious sense: as there is no obligation to maximise moral worth, there is no need to maximise instances (tokens) of action from duty either.[2] Not every single act need be motivated by duty alone – which would leave no actions to fall in the category of the morally permissible. Kant endorses the commonsense position that there are numerous cases in which acting on inclination is entirely legitimate.[3]

If so, permissible action motivated by inclination must somehow be composite. A rare moral merchant may ask a decorator to refurbish her shop to attract more customers. She acts on sound commercial interest, on a higher-order inclination, not simply because she has come to dislike her surroundings, and

1 The case of maxims is less complicated: they either pass or fail the test of the categorical imperative, i.e. they are either permissible or impermissible. The fact that maxims, rather than actions, are the fundamental concept of Kantian ethics solves many problems commonly thought to follow from the uncertainty of action description.

2 We have an indirect obligation to make sure we are aware of token duties – see Kant's famous exhortation not to avoid 'the places where the poor who lack the most basic necessities are to be found' in the *Metaphysics of Morals* (VI 457.29–30); but that should not be confused with a wish to maximise action from duty or moral worth.

3 In this matter I agree with Barbara Herman and Marcia Baron, who make this point in terms of 'primary' and 'secondary' moral motives. Kant does not wish to say that duty must always serve as one's 'primary' motive, or that we should try to maximise occasions on which it does. Duty can be present as a 'secondary' background motive and assume a 'regulative' role. See M. Baron, *Kantian Ethics almost without Apology* (Cornell University Press, 1995), esp. pp. 129–33.

certainly not from duty.[4] In this respect, she behaves just like her famous colleague in Section I (IV 397.21–32). The difference between the two is the following. The question whether to refurbish one's shop can legitimately be decided on commercial grounds, whereas the question whether or not to overcharge one's inexperienced customers cannot. Similarly, our moral shopkeeper has to observe common moral boundaries in her transactions with the decorator: for example, not lie to him about the urgency of the job or steal sandwiches from his lunchbox. This must be done from moral conviction. Permissible action from inclination need not itself be motivated by the moral law as long as impermissible acts are omitted from duty. The good shopkeeper's deed therefore lacks moral worth as an action while possessing moral worth as an omission. Of course, it would be difficult for the shopkeeper to refrain from, for example, lying to the decorator only if she was tempted to do otherwise.

Furthermore, and perhaps surprisingly, even morally obligatory action, which ought to be done from duty and moral conviction, can leave considerable discretion to judgement and even to inclination, particularly – but not exclusively – in the case of imperfect duties. If I do as duty commands by visiting an ailing friend in hospital I still have to decide whether to go there by bus or by taxi, or whether I should perhaps walk. I also face the choice as to whether I should get her flowers, a box of chocolates, a book or a bottle of wine – and, if so, which. All these matters depend on subjective, not moral considerations, and some choices are indeed more likeable, thoughtful or appropriate than others. If I judge it to be my duty to pay my tailor's bill, I have a choice of whether to write him a cheque or pay in cash. Convenience can quite properly determine that decision, as long as he receives his money in time. The fact that I must pay my bill within a fixed period is settled by legal and moral considerations. Again, the moral judgement that it is my turn to host a dinner for my students at the end of term does not affect the choice of food – which will be decided by inclination – my own as well as my anticipation of the preferences of my guests.[5] Consequently, the question whether I choose the wine, or even write a cheque, 'from duty' does not admit of an unequivocal answer. The broad end is determined by pure practical reason, but the means are not. Means are a moral matter only indirectly. Kantian ethics does not specify the minutiae of right action. In fact, any such 'micrology' makes complete nonsense of the idea of universalism in ethical theory.[6] Prudence, convenience and sympathy all have their proper role to play in the way we carry out our duty.

4 See Kant's classification of actions that accord with duty at IV 397.11–21.

5 If the example seems far-fetched, in his lectures Kant himself uses the example of hospitality as a case in which the moral and the pleasant may fruitfully be combined (*Collins*, XXVII 397). Also, as Onora O'Neill points out in reply to Alasdair MacIntyre's critique of moral universalism, making visitors welcome is a universally acceptable principle, whereas whether we offer them a cup of tea or a glass of wine is subject to cultural variation (Onora O'Neill, 'Kant after Virtue', in *Constructions of Reason* (Cambridge University Press, 1989), p. 151). A single, general maxim can co-ordinate different rules of skill, depending on the circumstances.

6 On Kant's rejection of 'micrology' see *Metaphysics of Morals*, VI 409.18.

Appendix C: Universal legislation, ends and puzzle maxims

Criticism of Kant's moral formalism and universalism is often motivated by the allegedly counter-intuitive implications of applying the categorical imperative to certain 'puzzle maxims'. A good example is that of dining at a friend's place on Monday nights.[1] A maxim along the lines of 'I want to dine at a friend's place at 7.00 pm on Mondays' cannot be universalised if we assume that the particular friend in question must be present, for example to discharge his or her responsibilities as the host of the party. But, surely, there is nothing wrong with dining at a friend's place on Monday nights? We have all done it. Our conscience was silent. Yet the universalisation test of the categorical imperative would seem to rule it out.

One possible way of dealing with this difficulty consists in demoting the said maxim to the level of a mere intention (*Absicht*).[2] According to this reconstruction of Kant's theory of action, maxims are 'life rules' (*Lebensregeln*) that render a life meaningful as a whole and are as such characterised by a certain generality. If so, dining with friends on Mondays is too particular and narrow a rule to qualify as a maxim – but if it is not a maxim it does not, so the argument goes, have to pass the universalisation test of the categorical imperative. This analysis contains a grain of truth, but the importance of maxims cannot be grounded in an ad hoc re-classification. It is also at odds with actual examples of maxims cited in Kant's writings on moral philosophy. For instance, should the suicide's maxim, to shorten his life 'when its longer duration is likely to bring more pain than satisfaction' (IV 422.5–8), count as a 'life rule'? It is not by definition that maxims as such are general but a *good* maxim is, precisely because it can be universalised.

The danger at this point is that of confusing 'is' and 'ought'. Highly specific principles, which might – depending on the situation – be put to good use as rules that stand under general maxims, ought not to be adopted as maxims, i.e. subjective practical principles; but that does not entail that they cannot be adopted as maxims at all or that adopting them would be morally innocent. Rules intended for occasional use may very well make bad practical principles.

1 This is R. Bittner's example, see 'Maximen', in *Akten des 4. Internationalen Kant-Kongresses*, ed. G. Funke (De Gruyter, 1974), vol. II. 2 p. 487. On puzzle maxims in general see my *Sittengesetz und Freiheit* (De Gruyter, 2003), pp. 159–73.

2 This is Bittner's preferred solution: see 'Maximen', p. 486.

Any maxim must be supported by a corresponding interest to be a viable principle of action. If so, someone who has developed a bizarre inclination to dine with his friends at 7.00 pm on Monday nights *as such and under that description* might indeed adopt and consequently act on such a principle. But he ought not to do so, precisely because his maxim would fail the test of the categorical imperative. A maxim is more than just an action-guiding rule; and doing something as a matter of principle, because one is directly interested in it as an end, is relevantly different, morally, from doing it merely as a means towards some other end. In fact, the example confirms the practical equivalence of the general formulation (and the law-of-nature variant) of the categorical imperative and the second variant, the formula of humanity as an end-in-itself. Dining with one's friends on Mondays *as a matter of principle* is morally suspicious because it suggests that one treats them merely as a means towards one's avowed end.[3] Would we really think of someone who consciously and regularly knocks on our door for his Monday meal as a good friend?

It is therefore important to observe the distinction between rejecting a maxim, which commends a certain act, and rejecting the act, which may spring from a variety of maxims. A regularity that makes a bad maxim can still serve as a good rule when put to use by another, more appropriate maxim. The act of dining at a friend's place by itself is perfectly unobjectionable, even morally valuable, if the ulterior end is that of cultivating one's friendships, and dining with friends on Mondays turns out to be a suitable means. But we should note that in that case the agent is committed to the said end, not to the particular rule that specifies the means. He would without the slightest regret try to be friendly with his friends in some other way if dining at a friend's place at 7.00 pm on Mondays turned out to be no longer feasible, for example because his friend has something more urgent to do.[4]

Franz Brentano's objection that the categorical imperative seems to rule out a maxim of rejecting bribes can be refuted in a similar fashion.[5] The worry is that, along the lines of the example of telling a lie or making a false promise, the maxim of a decent civil servant to turn down bribes, if universally adopted, would render itself impossible. No-one could hope to obtain a loan by a false

3 Also, the problematic maxim of dining with friends on Monday nights contains no indication that the agent himself is willing to reciprocate. By contrast, if his maxim is that of being friendly with one's friends, he certainly would be.

4 The same procedure should be applied to T. M. Scanlon's well-known co-ordination maxims of playing tennis when one's neighbours are in church on Sunday mornings, and of saving money by shopping in this year's after-Christmas sales for next year's presents (see Herman, 'Moral Deliberation and the Derivation of Duties', p. 138). Either course of action could be perfectly rational. If so, they follow from technical rules that stand under a general maxim of avoiding cues, or economic thrift, but their value as rules is purely instrumental. These rules do not deserve to be adopted as maxims because they do not specify actions that are worth doing for their own sake or as such. If shopping in this year's after-Christmas sales turns out to be an unsuitable means towards my end I will try to pursue it by different means. Similarly, few desire to play tennis on Sunday mornings because everyone else is in church.

5 See G. Patzig, 'Der Gedanke eines Kategorischen Imperativs', *Archiv für Philosophie* 6 (1956), 85–7.

promise if everyone always tried to borrow money in this fashion. Similarly, no sensible person would venture to offer a bribe if, as a matter of principle, bribes were universally rejected. Does this mean that turning down bribes is morally wrong because it undermines the institution of bribery? Certainly not. It is easy to see how this puzzle is to be resolved. The honest civil servant's maxim should not, strictly speaking, be that of rejecting bribes. The original maxim, if universalised in accordance with the categorical imperative, would indeed be self-defeating in the manner just described. However, as a moral person he has no interest in supporting the institution of bribery in any shape or form. 'Bribery' is therefore unlikely even to feature in his maxim. Turning down bribes is a means to a legitimate end, but it should not be considered worth doing for its own sake. Kantian morality turns out to be about achieving the right balance of practical means and ends. In fact, the categorical imperative seems to serve this purpose rather well.

The reply to Brentano's challenge is therefore the following. The civil servant's practical principle must be a general maxim of decency. In that case, he will be perfectly happy, and not at all disappointed, when bribes begin to disappear. This is precisely what he is trying to achieve. There is no paradox. Unlike the lying borrower, he is undermining an institution, *not* trying to exploit it.[6] In fact, adopting a maxim of turning down bribes for the pleasure of doing so smacks of self-righteousness and may well be considered morally suspicious. The categorical imperative correctly predicts this: a maxim of turning down bribes fails the test of the categorical imperative because it cannot even be thought a universal law.[7]

In Kant's moral philosophy, an action must be individuated with reference to the maxim that de facto generates it. This is very much a theoretical exercise because maxims are not discernible, empirically or otherwise, but the point to note is that there is no one-to-one correspondence between maxims and acts. A maxim reflects that in which we are directly interested.[8] That is why rules are flexible in a way that maxims are not – we are not committed to rules as such, we just resort to them as the situation commands. An agent's commitments will be revealed by his behaviour if circumstances vary. That is also the reason why we cannot modify a maxim at will when we realise that it does not conform to the categorical imperative, e.g. by adding morally insignificant information

6 A practical contradiction is generated not because a maxim would de facto undermine an institution if universalised, which would provoke the common Hegelian objection of 'empty formalism', but rather because the agent at the same time intends to make use of this very institution. As Kant puts it in the Typic of the second *Critique*, the agent must be a *part* of the system (of natural laws, V 69.22), not merely its originator.

7 The point of the fourth example is that it rules out an attitude of selfish indifference (IV 423.17–35, see IV 430.18–27). A morally worthy maxim of helpfulness would therefore seem to be one of sympathy and benevolence (IV 423.25–6), which if necessary translates into beneficent action, not of providing aid as such, which like that of turning down bribes would seem to lead to a contradiction by undermining the conditions which make assistance necessary.

8 The possibility of adopting any law in one's maxim depends on a suitable interest, see e.g. *Critique of Practical Reason*, V 79.24–5, V 90.36.

to make it more specific and thus – allegedly – universalisable. Someone who tells himself that perhaps a maxim of lying cannot be universalised while a maxim of lying to so-and-so in such-and-such a situation can, still intends to lie to achieve his purpose. His quibbling does not affect the maxim he intends to act on: he would have modified his 'maxim' differently if the situation had been different. In the lectures, moral self-conceit is said to arise when a man 'tinkers with the moral law, till he has fashioned it to suit his inclinations and convenience' (*Collins*, XXVII 465.1–3). The lying promiser is not just lying to others, but also to himself.

Appendix D: 'Indirect duty': Kantian consequentialism

Kant was not, and could not have been, a consequentialist.[1] In his ethical theory, moral interest is primarily directed at the action itself and not, like 'pathological' interest based on inclination, at the object that we intend to promote or effect.[2] However, it is a mistake to conclude that Kant must dismiss consequences as ethically irrelevant or unimportant. They matter in two distinct ways. First, effects have their role to play in the thought experiment of applying the more formalist versions of the categorical imperative. Effects do not determine what ought to be done – that *would* be consequentialism – but they help us to discover whether we can think or will a proposed maxim as a universal law. For example, when I am tempted to make a false promise I need to reflect on the consequences of my attitude's being adopted by all humanity to realise that I cannot will (in fact even think) a deceitful maxim as a universal law. Consequentialist calculation thus *helps* me to make up my mind whether I may act in the way I desire. But it does not *decide* the question of whether my action is permissible. That depends on whether I would incur a practical contradiction were I to act on the universalised maxim. It is hardly surprising that consequences should enter the moral equation in this manner. After all, the first variant of the categorical imperative tells us to think of our maxims as universal – causal! – laws of nature. Our choices determine the shape that nature takes. In the Typic of the *Critique of Practical Reason*, Kant expressly recommends this version for the purpose of practical decision making.

Secondly, we must calculate the consequences of possible actions once we know what, morally, we ought to do. The categorical imperative operates on maxims: fundamental practical attitudes or general policies that, particularly in cases of wide duty, leave the details of what ought to be done undecided. Doing

1 Several philosophers have tried to assimilate Kant's work to consequentialism in recent years: e.g. R. M. Hare, 'Could Kant Have Been a Utilitarian', *Utilitas* 5 (1993), 1–16, D. Cummiskey, *Kantian Consequentialism* (Oxford University Press, 1996), and S. Kagan, 'Kantianism for Consequentialists', in Immanuel Kant, *Groundwork*, trans. Allen Wood (Yale University Press, 2002). For a critical assessment of these views see Kerstein, *Kant's Search for the Supreme Principle of Morality*, pp. 139–55, and my 'Why Kant Could not Have Been a Utilitarian', *Utilitas* 17 (2005), 243–64.

2 Kant draws this distinction most explicitly in his footnotes at IV 413–14 and IV 459–60. In ruling out consequentialism I assume that reason does not and cannot command us to act for the sake of a law that commands us to produce good consequences; see IV 400.19–21.

one's duty competently is *indirectly* a matter of duty. But 'indirect' duties are not a distinct species of duties because they concern not the adoption of ends in one's maxims but merely the choice of means.[3] 'Indirect' duty is generated by hypothetical imperatives, which enjoin us to take the means to achieving any end to which we are committed, in particular: it is produced by the *technical* variety of hypothetical imperatives, dubbed 'rules of skill' in Section II. These rules teach us how *skilfully* to realise any end, and for this purpose draw on the agent's empirical knowledge of the workings of the world. The notion of 'indirect' duty thus helps to dispel fears that Kant wants us to do our duty irrespective of the consequences.

Moreover, like all hypothetical imperatives these rules command 'analytically' (IV 417.11, see IV 420 fn.). They do not, like the categorical imperative, require that we do something 'new', they merely advise us about realising an end to which we are already committed. For example, calling an ambulance is not something we do over and above helping the victim; on this occasion, helping the victim *consists in* calling an ambulance, which is indirectly commanded by duty (as first aid will be if the situation is marginally different). It may not be obvious from Kant's discussion of hypothetical imperatives in Section II – where his avowed purpose is first to distinguish conditional and unconditional imperatives – but instrumental reasoning is needed to realise both non-moral and moral ends in practice. In the latter case, the source of the end is pure practical reason.

Consider the case of furthering or securing one's own happiness, which Kant declares is a matter of duty 'indirectly' in the *Groundwork* (IV 399.3). Like modern moral consciousness generally, Kant suggests that moral integrity and personal happiness – defined as the satisfaction of the sum of one's desires – are separate spheres within a person's life; and that they ought to be kept separate. Morally good action may not make you happy, and action that makes you happy may not be morally good or even permissible. A moral character merely makes you *worthy* of being happy. Actions that can be predicted to make you happy are recommended by counsels of prudence, not by the moral law. Of course, the categorical imperative tells you to develop your talents, which may contribute to your happiness, but it does not as such command that you take a prudential interest in your long-term well-being. You incur no rational contradiction in the will or in conception if you fail to do so. However, taking the imperfections of humanity into account we realise that, as Kant puts it, 'lack of contentment with one's condition, under pressure from many worries and amidst unsatisfied needs can easily become *a great temptation to transgress one's duty*' (IV 399.4–6). That is why making sure that you are happy is *indirectly* a good thing, morally, after all. But its keeping you out of temptation is the one *morally* good aspect of assuring your own happiness. Carelessly exposing oneself to temptation is, not implausibly, seen as a kind of vice. In other words: securing one's happiness is not morally relevant in its own right, whereas not

3 For a more extensive discussion of Kant's theory of 'indirect' duty see my 'Kant on Conscience, "Indirect" Duty and Moral Error', *International Philosophical Quarterly* 46 (2006), 293–308.

leading oneself into the kind of temptation to which one is likely to succumb is. The two merely happen to coincide.

The same applies to the 'indirect' duty to acquire wealth (cf. *Metaphysics of Morals*, VI 388.26–30), which is actually implied by the more general 'indirect' duty to further one's happiness. (Human beings require some wealth to realise their inclinations, i.e. to be happy.) Kant states that the factors which check the influence of certain temptations – prosperity, strength, health and well-being in general – might by some be considered ends that are duties, 'so that one has a duty to promote *one's own* happiness, and not just the happiness of others' (VI 388.20–22). But they are mistaken. Everyone has a duty to look after his or her own well-being, not as an end itself but rather as a – for the most part – permissible means to a moral end: the subject's moral integrity. The acquisition of prosperity – like removing other obstacles to moral action – is declared to be 'not as such a duty', but it may be a duty 'indirectly'. It is then 'not my own happiness, but rather the preservation of my own moral integrity [*Sittlichkeit*] that is my end and at the same time my duty' (VI 388.28–30).

That is why Kant warns that certain duties to the self must not be mistaken for duties towards other creatures or persons – animals and God – which for different reasons cannot be the objects of our duties. There is an 'amphiboly', a deceptive ambiguity, that makes it look as if, for instance, we have duties towards animals whereas in fact there is no such thing as 'a duty to treat animals decently', but rather a duty to care for one's moral character which, according to Kant, also has surprisingly far-reaching implications concerning our treatment of animals (VI 442–4). We may disagree with Kant's account of the moral value of animal welfare, or rather the lack thereof.[4] But the general message is sufficiently clear: we must not confuse matters of so-called 'indirect' duty with duty *sui generis*, with duty proper. The definition of duty as 'the necessity of an action from reverence for the law' (IV 440.18–19) reveals a further reason why caring for one's happiness, as described above, is not a duty proper, but rather commanded by duty 'indirectly'. Any of these actions are per se morally neutral acts because they are not immediately made *necessary* by the moral law. They are obligatory only by accident because they help us to realise our good ends.[5]

4 On the moral status of animals within Kantian ethics see my 'When the Tail Wags the Dog', *Kantian Review* 10 (2005), 128–49.

5 The relevant description of an action, for Kant, must refer to the end pursued with it, which is contained in the action's maxim; and maxims and ends must be subject to the formal moral standard of the categorical imperative. In the lectures on ethics, Kant quite generally affirms that a thing is named after the end, not the means (*Collins*, XXVII 249.34–5). The illusion that the objects of 'indirect' duty are themselves morally valuable is dispelled by the consideration that one could, perhaps by sheer strength of will, remain a moral person even without performing actions in accordance with any of the 'indirect' duties mentioned above; there is no necessary connection between the two. By contrast, one cannot defy imperatives not to lie, to steal or to refuse to help others without necessarily and immediately compromising one's moral standing. Thus an action that is commanded 'indirectly' must always be described with reference to the obligatory end pursued or realised. Any other description runs the risk of creating the deceptive appearance of 'a duty' that is in fact indirect and must be reduced to quite another duty.

Appendix E: Freedom and moral failure: Reinhold and Sidgwick

One of the *Groundwork*'s most controversial claims is Kant's equation of a 'free will' with a will that stands 'under moral laws' (IV 447.6–7). Moral commands restrict the range of options we can freely choose. If we accept the Kantian thesis that moral laws are laws of practical reason, why are they also meant to be laws of freedom? What is worse, Kant's equation seems to defeat rather than secure the conditions of moral accountability. It is a philosophical commonplace that freedom is both a necessary and a sufficient condition of moral responsibility. If morally bad acts are not an expression of freedom, do we really deserve to be blamed for them? They appear to be determined by the laws of nature and, as such, not free at all; human beings might seem to be responsible for morally good actions only.

This objection is commonly associated with Carl Leonhard Reinhold. In his 1792 *Letters on Kant's Philosophy*, he accuses Kant of confusing the law-giving activity of practical reason with freedom of the will, leading to the 'impossibility of freedom for all immoral actions'. Reinhold agrees with Kant's characterisation of freedom as independence of natural determination, but he considers it one-sided and incomplete. A free will must also be independent of being necessitated by practical reason. Freedom should therefore be defined as the 'capacity of self-determination by means of choice, for or against the moral law'.[1] A very similar objection was put forward, in an issue of *Mind* in 1888, by Henry Sidgwick, who argues that Kant employs the word 'freedom' in two distinct and incompatible senses, without being conscious of a variation in meaning. 'Rational' freedom consists in getting things right. 'Neutral' freedom means having a choice. Kant is urged to abandon one of the two conceptions.[2] Sidgwick's preferred solution is to jettison 'rational' freedom. He argues that we should not give up on 'neutral' freedom because freedom of choice is the precondition of responsibility.

Unfortunately, Reinhold and Sidgwick grab the wrong horn of the dilemma. Kant's official reply to Reinhold can be found in the *Metaphysics of Morals* of

1 C. L. Reinhold, 'Erörterung des Begriffs von der Freiheit des Willens', pp. 255–6; reprinted in Bittner and Cramer, *Materialien zu Kants 'Kritik der praktischen Vernunft'*, pp. 252–74. See also Allison, *Kant's Theory of Freedom*, pp. 133–5.

2 The paper was reprinted as an appendix in later editions of Henry Sidgwick's *The Methods of Ethics* (Macmillan, 1874), pp. 511–16.

1797. He explains that freedom of *Willkür* cannot be defined as the choice for or against the moral law of reason, but rather in the *capacity* to act as one ought to act (VI 226.12–227.9). He emphasises the fact that although we experience freedom as independence we do not need a 'neutral' capacity always to choose, or to act, otherwise to sustain our moral intuitions. In a very similar manner, he states in a handwritten note from the late 1770s that one cannot 'say that the opposite of all our actions must be subjectively possible for us to be free . . . but only of those coming from our sensibility' (R 5619, XVIII 258). Moral responsibility rather presupposes an 'asymmetrical' capacity to leave undone the wrong thing when required, and to do the right. Reinhold's and Sidgwick's neutral conception – traditionally called 'freedom of indifference' – is simply an over-generalisation.

For example, it would be absurd to accuse a thoroughly honest man who is incapable of lying of a lack of freedom. As Kant says in the lectures, this person 'refrains of his own will'. He still can – and in fact does – do what he ought to do. Kant concludes that actions can 'be necessary without conflicting with freedom' (*Collins*, XXVII 267.37–9). Of course, he must do the right thing for the right reason, and not just habitually or mechanically. If so, there is nothing to be said for the so-called freedom to turn one's back on the commands of practical reason, even if as a matter of fact human beings do possess this possibility.

A free will is a will that can follow rational laws.[3] The human will is free in this sense, but as human beings we nonetheless often have difficulties realising our potential for rational action. We are not as unimpressed by improper influences as we would like to be.[4] Kant occasionally acknowledges human frailty by saying that, while the will always 'stands under' morality or freedom, only some actions proceed 'in virtue of', 'from' or 'through' one or the other. All human actions proceed from a free will that is subject to moral laws that it *can* conform to. Some actions are free in the sense that they *actually* realise their full potential, but all actions are, as one might put it, sufficiently free for the agent to be responsible for his or her actions. As Kant puts it in a pithy note from around 1776–8: 'Actions from inclination, if it was also possible to act from freedom, are free as well' (R 6931, XIX 209). Similarly, in the *Critique of Practical Reason*, Kant says that the human will is a will that is 'subject to pathological affections, though not determined by them', and therefore 'still free' (V 32.26–7). Again, in another little note from the 1770s: 'Bad actions are subject to freedom, but

3 S. Wolf defends a similarly asymmetrical view in her *Freedom within Reason* (Oxford University Press, 1990), but contrasts her 'reason view' with what, unfortunately from a Kantian point of view, she calls the 'real self view' and the 'autonomy view'.

4 A will like the human will is not perfectly free like the holy will of God, traditionally conceived, which is not even subject to any non-rational affections, feelings or desires. God is free from any volitions that might be in conflict with what is best. Whether we believe in such a God or not, a perfect will can serve as a useful model for a completely different kind of will in philosophical discussions. His will possesses freedom, positive and negative, to a much greater extent than we do with all our alternative possibilities of action; but alternate possibilities are open only to wills in an intermediate position between the brute natural will of animals and the perfect will of, perhaps, God.

they do not proceed from it' (R 3868, XVII 318). Kant's conception of human freedom as an asymmetrical capacity remains constant throughout the 1770s, 1780s and 1790s. He did not invent it in reply to Reinhold's attack.

Yet Kant's asymmetrical conception of freedom, if intuitively plausible, has its setbacks. First, freedom as a mere capacity is at odds with physical determinism. Of course, Kant's transcendental idealism opens up the possibility that physical determination is not the ultimate cause of action, but rather the surface effect of regularities we freely choose: the maxims of our actions. Physical determination, even a closed causal system, does not as such make us feel uneasy, but it does if our willing can go astray. The prediction that we will do the right thing does not make us feel uneasy, but we do object to the thought that our morally bad actions can in principle be predicted, as Kant puts it in the *Critique of Practical Reason*, with the same kind of certainty as a lunar or solar eclipse (V 99.17–18). If so, it is difficult to see how we could have exercised our asymmetrical freedom to do the moral thing. Ultimately, even the most charitable interpretation of Kant's attempt to reconcile the closed causal system of natural determinism and free will is bound to fail. Kant should probably weaken his conception of causality to accommodate some room for variation.[5] He needs to reconcile not determinism and freedom, which can easily be done, but nature as a closed deterministic system and human freedom as a mere capacity.

Secondly, Kant's asymmetrical conception of freedom faces an objection even more serious than the threat of physical determination: that of inconsistency. It would seem that the intuitive notion of a will that is both – positively and negatively – *free* and, at the same time, *weak* cannot be sustained. How can a free will be limited if not by restricting the range of choices, i.e. the kind of restriction we have to avoid? The right action must always be a real option, even if it is subjectively difficult for the agent to decide in its favour. How can the human will be merely 'sufficiently free' to warrant moral responsibility? If freedom is a mere capacity, we cannot account for moral failure, which amounts to the free renunciation of freedom (see R 3856, XVII 314). To rephrase the objection in more Kantian terms: moral failure cannot just be an effect of natural causation, as the agent is negatively free from that; but it cannot entirely be due to rational causation either, as after all immoral actions are not rational. We do not understand how the elective will, which after all is not subject to the causal laws of nature, can be 'affected' by sensibility at all. Sometimes, reason mysteriously fails to be active, or as active as it ought to be. Maybe Reinhold and Sidgwick have a point after all.

5 It is true, Kant tries to make sure that all of our actions are in accordance with some law or other, if not with the law that reason gives to itself then at least, as always, with natural causation. Which law our actions conform with, whether it is just the law of nature or the moral law as well, cannot be itself determined. There has to be some looseness to allow for moral failure, which in itself cannot be explained. Why should we be responsible for something that is inexplicable? Kant tried to avoid liberty of indifference for broadly Humean reasons, but in the end seems to be left with not just the good but also the bad things of both the compatibilist and incompatibilist worlds.

Throughout his later work – most notably in *Religion within the Limits of Reason Alone* – Kant concedes that moral failure is inexplicable (cf. VI 43.12–17). Much earlier than that, he admits with equal candour that, ultimately, the true nature of the moral ought facing the human will cannot be explained – e.g. on the very last page on the *Groundwork* (IV 463.29–33). Once again, the categorical ought and the capacity for free action represent two sides of the same medal of morality. In the *Critique of Pure Reason* Kant explicitly concedes that there is no answer to the question of 'why reason has not determined the phenomena differently'. Again, his example is that of a lie:

> [A]nother intelligible character would have produced another empirical one; and when we say that regardless of the entire course of life he has led up to that point, the culprit could still have refrained from the lie, then this signifies only that it [the lie] is subject to the power of reason, and that in its causality reason is not subject to any conditions of appearance or the temporal series. (A 556/B 584)

Temporal things do not determine reason, and the question why reason was not 'practical' to prevent the lie, or any other immoral action, must then be left open. Unfortunately, while a frank admission of failure makes its author more likeable it does not save his theory.

Appendix F: The project of a 'metaphysics of morals'

If the *Groundwork* is a foundational essay on moral philosophy, a preparatory text preceding a 'Metaphysics of Morals' proper and largely not even a preliminary sketch of this novel discipline, what would Kant's final ethical system look like? The answer is anything but straightforward. Kant had been planning to write a foundational treatise on moral philosophy since the mid-1760s, when his views were influenced by moral sense philosophy. He did not publish a 'Metaphysics of Morals' at the time, but it would undoubtedly have reflected his philosophical commitments. The project survives the changes of the 1770s and becomes part of the critical enterprise. For a while, there is no indication that a critical foundation of the metaphysics of morals other than the *Critique of Pure Reason* itself is required or even possible, but around 1783 Kant came to think that it needed some critical preparation after all – hence the *Groundwork*. The *Critique of Practical Reason* – which is curiously detached from the metaphysical project – followed in 1788. After finishing the third and final *Critique* Kant returned to the 'Metaphysics of Morals' in the early 1790s. He lectured on the topic in the winter semester of 1793–4, and a book called the *Metaphysics of Morals* was finally published, in two parts, in 1797.

It seems that for more than thirty years the title of a 'Metaphysics of Morals' was little more than a placeholder for a future systematic treatise on moral philosophy, while Kant's conception of the discipline, its foundations and its place within philosophy as a whole changed dramatically. If so, it is hardly surprising that the picture sketched of the moral metaphysics in the *Groundwork* is not always clear. There had been no time for Kant to consider the implications of the new critical grounding of moral philosophy in any detail. However, various clues are dispersed throughout the book. In the Preface, Kant declares that a metaphysics of morals will have to investigate 'the idea and principles of a possible pure will' (IV 390.4–5). The discipline rests on propositions that are pure and a priori as well as, for creatures like us, synthetic; hence the need to show that the principles of a pure will are valid for human beings in the partial 'Critique' of Section III. Section II effects the passage from populist ethical theories that fail to distinguish the pure and the empirically conditioned volitional elements to a metaphysics of morals (IV 406.2–4, IV 392.25–6). It would contain, we are told, a more detailed classification of duties than the fourfold division of strict and wide duties to the self and to others, which Kant adopts to

structure his examples throughout Section II (IV 421 fn.). Moreover, as morally valuable action must be done for the sake of a law that pure moral philosophy first formulates, distils and clarifies, a metaphysics of morals does not just satisfy the philosopher's curiosity but can and must be put to good practical use (IV 389.36–390.3). Kant claims that such a metaphysical system, unlike the *Groundwork*, is 'capable of a great degree of popularity and suitability for the common understanding', as opposed to the muddled popularity of the so-called popular philosophers targeted in Section II (IV 391.35–6, cf. IV 409.20–410.2). If so, what is the 'Metaphysics of Morals' of the *Groundwork* supposed to look like?

The structure of the *Groundwork* contains an important clue. Section II leads up to the notion that we should act as lawgivers in an ideal 'kingdom of ends', and concludes that these laws can only be the agent's very own laws. This is summed up in the concept of 'autonomy'. Kant proceeds to distinguish autonomous and heteronomous types of will and corresponding ethical theories. Section III begins with, and indeed relies on, this concept and justifies the supreme principle of morality precisely as a principle of autonomy. According to the tasks assigned to these parts of the *Groundwork*, Section II *leads to* a metaphysics of morals whereas Section III takes it as its *starting point*. It is natural to conclude that a metaphysics of morals would specifically be a theory of autonomy, i.e. self-legislation. After all, a metaphysics of morals investigates the laws of a pure will – which *are* the laws of autonomy. Metaphysics quite generally is the a priori cognition of objects and their laws. These laws each govern their own part of reality: the phenomenal and the intellectual realms. They are either the laws of nature or the laws of freedom.

The purity of a theory of autonomy also explains Kant's hope that a future 'Metaphysics of Morals' would have a benign influence on a wider readership. He has great faith in the motivational powers of the idea of autonomy and the cognate concept of an ideal commonwealth. In a note dating from the late 1770s Kant says that 'the kingdom of God on earth is an ideal, which has the power to move [*bewegende Kraft*] in the mind of one who wants to be morally good' (R 6904, XIX 201), which, if we trust the *Groundwork*, is each and every human being. Similarly, in the *Critique of Practical Reason* Kant argues that human reason must 'first work its way up . . . to gather strength to resist the inclinations by a lively representation of the dignity of the law' (V 147.16–18) – and the idea of a moral world inspires reverence in us (V 82.35). It makes us fully aware of our sublime existence as free and rational beings and thus helps us escape the wiles of inclination (V 88.23). The law of autonomy is 'the basic law of a supersensible nature and a pure world of understanding' (V 43.24–5). In his discussion of the project of a metaphysics of morals at the outset of Section II Kant similarly assigns to a metaphysics of morals the purpose of strengthening the rational incentive of reverence in our hearts by producing a clear representation of the purity and dignity of morality (IV 410.25–411.7).

Consequently, the role of a metaphysics of morals as the science that describes the law of autonomy is motivational and educational, rather than cognitive.

There is a neat (perhaps rather too neat) division of labour between the basic version of the categorical imperative, already formulated to complete satisfaction at the end of Section I, which tells us what to do (*principium diiudicationis*), and its popular variants that rest on certain analogies, formulated in Section II, which by bringing morality closer to the imagination secure 'access' or 'acceptance' (*Eingang*, IV 437.1–2) for the moral law to counteract moral corruption (*principium executionis*). After all, Kant thinks – problematically – that the categorical imperative can easily be applied by even the 'commonest understanding'. It is acting on one's insight that is difficult for beings like us; and *that* is where moral philosophy must come to the rescue of common man.

According to Kant's footnote at IV 421, the final 'Metaphysics of Morals' must also include a systematic exposition of different kinds of duty, i.e. like the book that was finally published in 1797, distinguish duties of right from duties of virtue and introduce the appropriate subdivisions within these broad categories. Unfortunately, this task does not seem to fit the description of a metaphysics of morals as an inspiring theory of autonomy that exposes the laws of a pure will as realised in an ideal commonwealth. A pure will is located in the world of understanding, in which there are no duties because the moral law applies descriptively. Only when we consider ourselves as members of both the world of understanding and the world of sensibility do we experience morality as an unconditional ought (IV 454.9–15). Kant presumably thought that the metaphysics of morals would not regard the kinds of obligation to be classified duties, in the thick sense of the word, with all the moral psychological details they entail – that must be left to 'anthropology' – but rather as the commands of an autonomous human will. They would no longer be considered synthetic practical principles – that they are valid as such has been shown in Section III. Yet a certain amount of 'impurity' would inevitably affect such a project. It is – comparatively – easy to see how a pure will could contain the moral law in itself; it is rather more difficult to see how it might contain particular token duties, such as the duty not to lie or to care for one's friends, which draw on the specifics of human nature; and a pure will certainly does not contain commands that are merely 'indirectly' a matter of duty.

Kant always felt uneasy about the status of moral theory within transcendental philosophy.[1] As Kant puts it in the revised Introduction of the *Critique of Pure Reason*:

> although the supreme principle of morality and the fundamental concepts of it are a priori cognitions, they still do not belong in transcendental philosophy, for, while they do not, to be sure, take the concepts of pleasure and displeasure, of desires and inclinations etc., which are all of empirical origin, as the ground of their precepts, they still must necessarily draw them into the formulation of the system of morality in the

1 Yet in the *Groundwork*, moral philosophy is said to admit of a special, rarefied kind of purity that exceeds even the purity of the principles of theoretical philosophy (see IV 411–12). Kant apparently failed to make up his mind about the respective purity of the two parts of philosophy.

concept of duty, as the obstacle that must be overcome or the enticement that ought not to be taken as a motivating ground. (B 28–9, cf. A 14–15)

Consequently, the section and chapter headings of the *Critique of Practical Reason* do not contain the epithet 'transcendental'. However, the pure and a priori nature of duty itself is not compromised by this element of impurity. We are well acquainted with the reason. Nothing that is merely empirical in origin is ever sufficient to establish any kind of command. Normativity requires an a priori addition. In other words, duty qua duty is due solely to the contribution reason can make to action.[2]

If this analysis is correct, the *Metaphysics of Morals* published in 1797 is not the book that Kant promises his readers in the Preface of the *Groundwork*. It does not reflect the views of the winter of 1784–5. Admittedly, under the heading of 'the idea and necessity of a metaphysics of morals', the general introduction to the later work contains much that seems very familiar. Kant sharply distinguishes moral theory from more empirical disciplines, and stresses the a priori and necessary character of its principles. Once again he emphasises that morality cannot be reduced to a doctrine of happiness. A practical philosophy concerned not with 'nature but the freedom of the faculty of choice' is said to presuppose a metaphysics of morals. To 'have' such a metaphysics is even declared a duty, 'and every human being also has it within himself, though commonly only in an obscure way' (VI 214–16). All human agents 'naturally' believe in a priori practical principles, which it is then a matter of duty to develop.

But in what follows, it is quite apparent that in the work of 1797 Kant at best intended to execute the more applied elements of the metaphysics of morals prepared by the *Groundwork*.[3] In fact, he seems to encroach even further on the territory of a moral anthropology by including within the scope of the metaphysics of morals certain principles of application (*Anwendung*):

Just as there must be principles in a metaphysics of nature for applying those absolutely highest universal principles in a metaphysics of nature to objects of experience, this is something a metaphysics of morals cannot dispense with,[4] and we shall often have

2 Things are more complicated even with regard to the purity of transcendental philosophy itself. In the second edition of the *Critique* Kant calls into question the pure status of causal judgements (B 2–3); see K. Cramer's *Nicht-reine synthetische Urteile a priori* (Carl Winter, 1985).

3 It is also worth remembering that the *Metaphysics of Morals* of 1797 is hardly a unified philosophical treatise. If the 'patchwork theory' applies to any of Kant's writings, it is this. The 'Metaphysical Elements of the Doctrine of Virtue' is often reminiscent of the lectures on moral philosophy that Kant gave in the late 1770s and early 1780s, i.e. material that pre-dates the *Groundwork*.

4 This appears to imply that mid-level principles of application are now part of a metaphysics of morals, which is also suggested by the text of the *Metaphysics of Morals* itself. Allen Wood is therefore right when, with reference to VI 217, he argues that the 'scope of a metaphysics of morals expands in the direction of the empirical, that of practical anthropology seems correspondingly to contract', and that according to Kant's final conception of the distinction between a metaphysics of morals and a moral anthropology the latter is mainly concerned with the subjective conditions that hinder or help them in fulfilling the commands of the former. See Allen Wood, 'The Final Form of Kant's Practical Philosophy', in M. Timmons, ed., *Kant's Metaphysics of Morals* (Oxford University Press, 2002), pp. 3–4. Alternatively, one might

to take as our object the particular *nature* of human beings, which is known only by experience, in order to *show* with reference to it what can be inferred from universal moral principles. (VI 216.34–217.4)

He hastens to add that this 'will in no way detract' from their purity or 'cast doubt on their a priori source'. The 'counterpart' (*Gegenstück*) of a metaphysics of morals, a moral or practical anthropology, then deals 'only with the subjective conditions in human nature that hinder people or help them in *fulfilling* the laws of a metaphysics of morals' (VI 217.11–13), a task also assigned to it in the *Groundwork* – just not the only one.

argue in the spirit of the *Groundwork* that even though the metaphysics of morals requires mid-level principles they are not to be included in the system but rather are part of applying the metaphysical principles to human psychology (anthropology). But if there are principles of application 'in' a metaphysics of nature they would also have to be included in the parallel metaphysics of morals, even if Kant does not expressly say so.

Glossary

Many misinterpretations of Kant's philosophy are due to his complicated, at times treacherous, use of language. There are two kinds of linguistic difficulty. First, Kant's obsession with terminology should not be confused with precision of expression, as the weary old joke amongst Kant scholars goes. He all too frequently employs even his own official technical terms ambiguously. This does not as such render his reasoning fallacious, but it does make it much harder for his readers to understand his arguments. Secondly, Kant's language can be similarly deceptive in cases in which he offers no explicit definition of a term but rather relies on the philosophical usage of his time, particularly in the field of moral psychology. The linguistic change that has taken place since Kant wrote the book and the veil of translation – in fact translations, with their competing idioms – cause further complications. As a result, the sense of some words we find in our printed editions and translations is almost exactly the opposite of what an innocent reader might suspect.

The following glossary is intended as a brief guide to some prominent themes of the *Groundwork*. It contains key terms of Kant's project that most stand in need of explanation; others (e.g. 'dialectic', 'intelligible', 'pathological', 'speculation') are explained in the running commentary as they first appear in Kant's text.

AUTONOMY AND HETERONOMY

An adequate understanding of Kant's conception of autonomy is complicated by the fact that the currently popular concepts of 'rational', 'individual' or even 'moral' autonomy claim Kantian ancestry, despite the fact that the connection is often remote. Consider the following characteristics of Kantian autonomy. It is primarily a particular *type of will* that possesses autonomy; then – and more explicitly in the second *Critique* – human (practical) reason, which of course is equated with the will in the *Groundwork* (IV 412.29–30); finally, by extension, a (kind of) rational being equipped with such a will (e.g. V 87.23). Moral theories can be classified with recourse to their autonomy or heteronomy, depending on which model of the human will they favour (IV 441–4, cf. V 39–41). But autonomy and heteronomy are never predicated of individual acts of

the will, actions or even maxims.[1] There is nothing personal about autonomy. The wills of all mature and sane human beings possess it; the law is the same; and there is not a jot of individualism in Kantian autonomy, which is precisely why autonomy demands that we consider ourselves *universal* legislators. Although we legislate the law to ourselves and are thus the authors of our own obligation, we are not the author of the law, which is a law of reason. Autonomy does not involve the arbitrariness or stubborn defiance one might associate with those who are said to be 'a law unto themselves' (see Romans 2, 14). Moral action is strictly identified with the use of autonomy (see e.g. IV 439.24–6). Kant does not use adjectives like 'autonomous' or 'heteronomous'.[2] A will either possesses or lacks autonomy. There is no third. In the later case it is – to use the convenient adjective – heteronomous, and hardly deserves to be called a 'will' (*voluntas*) at all. It is merely a faculty of choice (*arbitrium*), governed by natural laws. Moreover, Kantian autonomy does not admit of degrees. Human beings do, of course, fail to live up to the autonomy of their will when they act immorally, but their faculty of reason still issues its own laws that they then violate.

The concept of autonomy is introduced informally at IV 431.16–18, officially at IV 433.10, and is defined at IV 440.14–32 (cf. V 33.8–33). The literal meaning of 'autonomy' is 'self-legislation'. This means, first and foremost, that the will issues *its very own laws*. The laws of autonomy are determined not by influences alien to the will but solely by pure practical reason. The will does not have to go outside itself, to an external object, to find the law (see V 62.15). In this way, an autonomous will is free (see FREEDOM); and it is the will issuing its very own law that one reveres, in oneself as well as in others (see REVERENCE). In ethical theory, moral laws cannot be derived from social convention, political or religious authority, or nature. No law resting on inclination qualifies as a law of autonomy because natural desires too are external to the will as practical reason. They are not rooted in what Kant considers the true self of human beings. A will lacks autonomy if to be moved to act it depends on incentives that do not come from within. A heteronomous will depends on, and can be manipulated by, external influences.

Yet autonomy is self-legislation also in the sense that it applies to the self, one's own will. It is legislation by the will, to the will. Only the will as a whole can be said to possess autonomy. Within the will it is the legislative capacity that prescribes a law to the executive faculty of choice. The internal structure

1 Kant scholars have been less reluctant to speak about autonomous and heteronomous *action* than Kant himself. However, widening the conception of autonomy in this way causes unnecessary problems without any obvious philosophical gain. Should an action judged to be morally permissible but done from inclination be called autonomous or heteronomous? Of course, a finite will can 'fail to live up' to its autonomy when the agent decides to reject the law of pure reason and acts immorally. But that does not make the will as a faculty any 'less' autonomous.

2 Kant uses the word 'autonomisch' only in the *Nachlaß*; see e.g. XXI 107.11, XXII 447.09, XXIII 455.21, XXII 466.20. It was a neologism at the time of composition. It is therefore unlikely that until then Kant eschewed the adjective for philosophical reasons.

of the human volitional faculty can be illustrated with reference to the political concept of autonomy (see *Perpetual Peace*, VIII 346). In ethical theory and – alas – in political practice, autonomy does not imply democracy. Practical reason has won the war of independence against mother nature, but the constitution of the human will is anything but democratic. The decrees of the sovereign pure will (*Wille*) may well seem despotic to its subject (*Willkür*, for this distinction see VOLITION AND CHOICE).

ENDS AND PURPOSES

Kant tends to distinguish the 'purpose' (*Absicht*) of an action[3] from its 'end' (*Zweck*). The former is a *subjective representation* of what one intends to do, located in the agent's mind; the latter is the *object* at which one's choice is directed, for the sake of which one acts. This explains the otherwise curious identification of each 'rational nature' (*vernünftige Natur*) – each individual rational being as such – with an 'end-in-itself' (IV 428.2–3).

In Section I, we learn that the moral value of an action does not reside in its purpose. Rather, what makes an action good is the subjective principle of the agent that selects and co-ordinates individual purposes: the maxim of the action (see the 'second proposition', IV 399.35–400.3). In this sense, a purpose would seem to be that which is willed as a means, in accordance with a specific rule (see MAXIMS). In Section II, Kant tries to reformulate the categorical imperative, which so far appears to be purely formal, with reference to the matter of moral volition: its end. For any imperative to be unconditional there must be an object of volition that is an 'objective end' or 'end-in-itself', solely by virtue of what it is, independently of human likes and dislikes, which otherwise determine whether or not an agent regards an object as an end: a 'subjective' or 'relative' end (see also IV 437.23–30). The 'end-in-itself' turns out to be man (*der Mensch*), or any rational being like him. Every such creature, oneself as well as others, must always be treated as an end and never merely as a means.

Elsewhere, the subjective conception seems to be the standard view. An end is defined as 'an object of the choice (of a rational being), through the representation of which choice is determined to an action to bring this object about' in the *Metaphysics of Morals* (VI 381.4–6, cf. third *Critique*, V 219.31–220.4). Kant proceeds to develop his theory of ends that are at the same time duties: one's own perfection and the happiness of others. The theory of rational beings as 'objective ends' retains an exceptional place within Kant's overall framework.

3 Variously rendered 'purpose', 'intention' or 'objective'. In his teleological excursus (IV 394–6), Kant even speaks of the 'purpose' that Nature had in mind when she endowed us with practical reason: morally good volition, not happiness.

FREEDOM OF WILL

It is obvious that the concept of freedom (*Freiheit*) plays a central role in Kant's ethical theory. Moral laws are called 'laws of freedom' on the very first page of the *Groundwork* (IV 387.14–15); and at the beginning of Section III, Kant declares a free will and a will under moral laws to be 'one and the same' (IV 447.6–7). However, the precise meaning of the word is often less than obvious. We need to distinguish four connected but distinct conceptions. The first concept is 'negative' in that it merely tells us what freedom of will is not. The remaining three are 'positive': (i) Freedom as independence. A will is a kind of causality because its activity produces effects. It can be said to be free first of all if it is not determined to act – and hence to produce certain results – by the forces of nature or 'sensibility', be they physical or psychological. A will whose actions are determined by natural laws is not free.[4] However, a negative characterisation of the acts of a free will is not only uninformative (IV 446.14), it is difficult to see how a will that is merely free in this sense can do anything at all; and little would be gained if it could. As Kant puts it in the second *Critique*, the mere absence of natural determination leaves the will exposed to be governed by blind chance (V 95.14). (ii) Freedom as spontaneity. The thought that free action is independent of *alien* determination quickly leads to the idea that it is caused by the agent's will, which itself initiates a subsequent chain of natural effects and causes. Kant calls this kind of freedom (absolute) 'spontaneity' (*Spontaneität, Selbsttätigkeit*). This bare conception of positive freedom of the will is prominent in the first *Critique*. Spontaneity is equated with transcendental freedom at A 533/B 561–2 (see also the second *Critique*, V 48.20–3, V 101.11). (iii) Freedom as (pure) rationality. The will of a rational being is defined as practical reason (*praktische Vernunft*) at IV 412.28–30; and a will that can operate completely independently of sensibility is free. If so, it should hardly come as a surprise that practical freedom is also equated with the capacity to act in accordance with rational standards, i.e. imperatives. In an early note Kant says that 'freedom is really the capacity to subordinate all voluntary actions (*willkürliche Handlungen*) to reason' (R 3865, XVII 317). (iv) Freedom as autonomy. Unlike the previous two positive conceptions, the identification of freedom with autonomy emphasises the idea that a will, by virtue of its being a kind of causality, must be governed by laws (IV 446.15–18). The transition from negative freedom to autonomy is elegantly formulated in the following note:

> Imagine freedom, or a faculty of choice, that is totally independent of instincts and the direction of nature generally; this in itself would be a kind of irregularity and the origin of all evil and disorder – if it cannot be a law unto itself. Freedom must therefore be subject to the condition of universal regularity, it must be an understanding kind of freedom. Otherwise it is blind or wild. (R 7220, XIX 289)[5]

4 To be precise, Kant thinks that even though actions *appear* to be thus determined their regularity must ultimately be due to a kind of causation that is not.

5 This is a recurring theme. In R 6961 Kant states that without the guidance of morality 'folly and chance govern the fate of men' (XIX 215); see also *Collins*, XXVII 258.1–6.

The second half of the quotation reveals what Kant's reply to the challenge of negative freedom looks like. The will is independent of 'alien' laws of nature. But it cannot be lawless, and must thus be autonomous, a law to itself. If the will's own law is the purely formal moral law, it makes good sense to say that a free will and a will under moral laws are one and the same (see IV 446.24–447.7 and AUTONOMY AND HETERONOMY above).

This fourfold characterisation of freedom has some intriguing consequences. We have learnt that free actions are governed by their own special laws, the laws of reason. In the case of the human will, which is weak and can fail to enact rational standards, this introduces a certain asymmetry. For Kant, freedom does not consist in the choice for or against rational action, but rather in the *capacity* to act rationally (see *Metaphysics of Morals*, VI 226.12–227.9). If so, freedom cannot be equated with libertarian freedom of 'doing otherwise'. Quite to the contrary, it is realised to the full when the will in question is perfect like the will of God. A note from around 1783–84 illustrates this:

> The freedom of the divine will does not mean that He could have done something other than the best (for this is not even what human freedom means), but rather that He is necessarily determined by the idea of what is best; which is not so with man, and that is why his freedom is limited. (R 6078, XVIII 443)

Our freedom is limited *because* our actions are not completely determined by reason. *Negative* freedom is independence of external determination. *Positive* freedom is ultimately the capacity to do what is, all things considered, best, and therefore, if we consider moral considerations overriding, what is moral in all morally relevant cases. If an agent's actions are determined by reason, he or she will not wish for the opportunity to do otherwise. Freedom is not opposed to law-like determination as such, nor even to subjective or objective necessity, but rather to being determined by the wrong kind of force. To be responsible, we must be able to do otherwise *only when we do wrong*, i.e. we must be able to do the right thing. Also, free choice does not secure responsibility if the right option is not available. If some evil demon confronts you with a free choice amongst several equally unpalatable courses of action, e.g. of either strangling, poisoning or shooting your best friend, we will hardly blame you or put you on trial for murder if you decide to shoot her – not because you could not have done otherwise, but because you could not have chosen morally.

There are parallels in the theoretical sphere. In his essay on *Orienting Oneself in Thinking*, Kant declares freedom in thinking to be 'the subjection of reason under no other laws than those it gives itself. Its opposite is the maxim of a lawless use of reason' (VIII 145.6–8); and if the will does not subject itself to the law it gives itself, it will have to bow under the yoke of laws that others impose. Kantian freedom, as such, has nothing to do with having a choice of different options – it is rather a sign of the limitations of human beings that they face a choice between reason and inclination (cf. Kant's use of the simile of Hercules at the crossroads, IV 400.12). Note that these choices do not merely concern single actions. They concern the implicit principles that determine our actions: maxims. Maxims are the *locus* of – negative, and in the light of the

laws of the will also positive – freedom, chosen by *Willkür*, and thus the proper object of moral evaluation (see MAXIMS, VOLITION).

HAPPINESS

The need to separate the demands of happiness (*Glückseligkeit*) and morality is a constantly recurring theme in Sections I and II of the *Groundwork*. Eudaemonism in ethics makes all human action depend on pre-existing inclination towards a certain object and therefore cannot serve as the foundation of an unconditional moral command. Our natural desire to be happy can neither inform the criterion of morality nor serve as our incentive to act morally. Moral philosophy does not teach us how to be happy, but rather how to be *worthy* of happiness, irrespective of our actual well-being (IV 393.15–24, see IV 450.7).

Kant's opposition to eudaemonism depends on his experientialist conception of happiness, which is usually defined as the – presumably conscious – satisfaction of the sum of all inclinations (IV 394.17–18, see IV 399.9–10, IV 405.7–8 and *Critique of Practical Reason*, V 22.19). It is identified with 'complete well-being [*Wohlbefinden*] and satisfaction with one's condition' (IV 393.20–1), with a person's 'preservation and welfare [*Wohlergehen*]' (IV 395.8–9).[6] In the first *Critique* Kant emphasises three dimensions of happiness: we wish our desires to be satisfied 'extensively, with regard to their manifold nature, as well as intensively, with regard to their degree, and also protensively, with regard to duration' (A 806/B 834, cf. *Metaphysics of Morals*, VI 387.26–8). Happiness is therefore not just a momentary state, but – in this respect like Aristotelian εὐδαιμονία – a sense of contentment throughout one's life. In terms of the objects pursued, happiness consists in the realisation of all ends as given by sensibility. Consequently, only 'finite' beings with sensuous needs can be said to be happy.

As all human beings naturally want their inclinations to be satisfied, everyone naturally wants to be happy. We are even – within the bounds of the morally permissible – rationally committed to furthering our natural ends, i.e. overall happiness is a final end all human beings possess. The conditional (hypothetical) imperative that tells us how to realise this end is called 'assertoric' (IV 415.1). The quality associated with this kind of pursuit is 'prudence' (*Klugheit*); the associated adjective is 'pragmatic' (*pragmatisch*). Note that in contrast to the pure and a priori principle of morality experience is required to discover such imperatives because, as Kant puts it in the *Metaphysics of Morals*, 'only experience can teach us what brings us joy' (VI 215.30–1). The happiness in question is always the agent's own, which as such is morally neutral and morally relevant at best in an 'indirect' fashion (IV 399.3). By contrast, there is a moral duty to be beneficent towards others, i.e. to help them in the pursuit of realising their ends, to make them happy (IV 398, see IV 423, IV 430).

6 However, Kant occasionally uses a less relativist conception of a happiness that merits rational pursuit (see e.g. IV 399.24).

In the *Critique of Practical Reason*, Kant distinguishes several objectionable kinds of self-love (*Selbstliebe*) from the legitimate pursuit of happiness. They are to be put in their place by our moral consciousness (V 73.10–24). There may also be a certain distinctly moral satisfaction, which must be distinguished from the sense of gratification that results when one's natural ends are realised. Perhaps as a result of the new aesthetic theory of the *Critique of Judgement*, this idea is more prominent in Kant's later writings, e.g. *Religion within the Limits of Reason Alone* and the *Metaphysics of Morals*. Yet he remains ambivalent about the term 'moral happiness' (VI 67.20, but see VI 387.30). In the second *Critique*, contentment felt by those who act on moral maxims is explained with reference to the independence from inclination gained when pure practical reason is in charge (V 117–18).

MAXIMS

Everything that happens conforms to principles – e.g. physical principles of motion. It is characteristic of human beings that they conceive of themselves as freely choosing the principles that govern their (conscious) behaviour, i.e. as capable of making their maxims conform to objective standards of reason, even though they often fail to do so (see FREEDOM OF WILL). Maxims are officially defined as 'the subjective principle of volition' and hence, given that volition causes action, the 'subjective principle of acting' at IV 401 fn. and IV 420 fn. respectively. In both passages, maxims are contrasted with objective principles or laws, which address a finite volitional faculty like ours as imperatives. Whether an action has moral *value* depends on the moral *content* of the maxim it springs from (IV 397.36–398.1, IV 398.7–19).

First and foremost, the notion of a maxim therefore occupies the 'is side' of the is/ought divide. Maxims are the principles on which human beings do in fact act. However, maxims have the status of philosophical constructions: maxims are the principles that must underlie human action for a morality of a categorical imperative as the law of human freedom to make sense. All actions properly so called – not reflexes or nervous tics – rest on maxims. Kant explicitly states what Henry Allison has called the 'incorporation thesis' in his *Religion*: 'Freedom of the elective will [*Willkür*] is of a wholly peculiar quality in that it cannot be determined to an action by an incentive *unless the agent* [*der Mensch*] *has incorporated it into his maxim*' (VI 23.3–24.3). Kant's categorical imperative commands *only* to act on that maxim, amongst those that our interests make available, that one can will to be a universal law.[7]

7 In addition, Kant very occasionally speaks of maxims as particularly robust subjective principles or life rules, a notion that has gained prominence in the wake of Rüdiger Bittner's work. For instance, the lectures on moral philosophy indicate that it is worse 'to do wrong from maxims than from inclination' and that 'good actions must be done from maxims' (*Collins*, XXVII 368.32–3). This usage is Kantian, but clearly secondary. Kant should not be seen to be saying that action from inclination does not involve maxims (in the standard, descriptive

Faced with a choice of options,[8] we have a sense of what kind of principle taking these options would implicitly commit us to (see e.g. IV 422.20–1); but we do not, in retrospect, know whether we acted on a moral maxim (IV 407.1–4); nor can we explain the possibility of free action (IV 458.36–459.2). (If a maxim is defined as the principle of action freely adopted by a finite will, we are not in a position to know whether we ever act on maxims at all.) In other words, as maxims define a person's character – the 'distinctive constitution' of the will (IV 393.12–13) – the moral character of agents, of oneself as well as others, essentially remains obscure.

It is important to note that not just any rule that describes an action qualifies as its maxim. Rather, maxims define an agent's commitments, that which an agent values as an end, and thus form the agent's 'character' (*Charakter*) or 'fundamental attitude', 'mindset', 'dispositions' (*Gesinnungen*, IV 435.15). By contrast, mere rules – called 'problematical imperatives' or 'rules of skill' (*Regeln der Geschicklichkeit*) in the *Groundwork* – teach us how to realise a given end. Maxims can thus co-ordinate several subordinate rules (*Critique of Practical Reason*, V 19.7–8). The correct description of an action must refer to its maxim.

MOTIVATION

Historical accident has led to endemic confusion amongst English-speaking Kantians. Following Paton, *Triebfeder* – which Kant glosses *elater animi* (e.g. V 72.1) – is now commonly translated 'incentive', whereas *Bewegungsgrund* – in accordance with the Latin *motivum* – is rendered 'motive'. The two are explicitly distinguished at IV 427.26–7. Philosophically, it would have been preferable to defy etymology and reverse the choice of words. *Triebfeder* is a mechanical term. It designates the 'spring' of motion, as in a clock or an old-fashioned toy. In psychology, it is by extension a motivating desire, the force that propels an agent forward if he or she so chooses. A *Triebfeder* 'makes the will practical', i.e. makes action possible (*Mrongovius* II, XXIX 625.37, cf. Kant's definition of interest, IV 459–60 fn.). Reverence for the law (*Achtung*) – the desire to do what is morally required (IV 440.6–7) – and inclination (*Neigung*) are competing *Triebfedern*. The motive of a rational being endowed with a

sense) – any such alleged action would be in breach of the incorporation thesis. Rather, as we are capable of freely choosing our fundamental practical attitudes, we are *also* capable of making them particularly robust, characteristic and consistent; which is manifestly a good thing if they conform to moral commands, and a bad thing if they do not. Relatedly, the categorical imperative does not command or imply that we *should* act on maxims or 'life rules'; rather that amongst the maxims available we should choose that which we can will to be universally valid.

8 As the incorporation thesis indicates, maxims determine which of the available incentives are expressed in action; maxims to let the will (as a whole) be determined by an incentive. Our options in each given situation – of maxims as well as actions – are limited by the incentives available.

will is also called an 'interest' (*Interesse*); see IV 459–60 fn. By contrast, a *Bewegungsgrund* is something static. It is the object that prompts the mechanism of motivation to action. In English, 'motive' can perhaps have both meanings, but in philosophy the former is prevalent.[9]

Consider the following example. It is natural to say that the exorbitant reward promised was the agent's *incentive*, and that his hunting down the criminal was *motivated* by greed. By contrast, Kant's translators make him say that greed is the 'incentive' and money the 'motive'.[10] This is a terminological fiasco. It seems preferable to render *Bewegungsgrund* 'motivating ground'. 'Incentive' can then be employed as a term of art alongside 'motive' to translate *Triebfeder*. 'Ground of determination' should be used for the *Bestimmungsgrund* of the will, i.e. that which contains the law in accordance with which the will acts.

NECESSITY AND NECESSITATION

Necessity (*Notwendigkeit*) is an objective term, referring – in the practical realm – to a moral ought. Correspondingly, action contrary to duty is morally impossible; and an action neither commanded nor prohibited is morally possible. Moral necessity can only be rooted in reason, not in experience. By contrast, necessitation (*Nötigung*) is a psychological term, indicating cases of action from duty in which the human will must subdue natural desires and force itself to act contrary to what it is inclined to do. The two should not be confused. Thus when in Section I duty is defined as the *necessity* of an action from reverence for the law (IV 400.18–19), Kant is not saying that action from duty must always be burdensome, subjectively difficult or painful. Rather, he means that the morally right action must be done, and is necessary in this sense, solely from a certain attitude of conforming with moral laws, independently of one's inclination, whether friendly or hostile.

REVERENCE

Reverence for the law (*Achtung fürs Gesetz*) is the interest in doing the moral thing without any ulterior purpose and thus the one and only motive of morally

9 The OED defines 'motive' in the relevant sense as 'that which "moves" or induces a person to act in a certain way', glossed as either 'a desire, fear, or other emotion' (Kant's *Triebfeder*); and adds that the word is also often applied to 'a contemplated result or object the desire of which tends to influence human volition' (Kant's *Bewegungsgrund*). (There is also an intermediate possibility that corresponds to Kant's *Absicht*: purpose or intention.) Interestingly, the Duden dictionary of the German language uses the two Kantian terms *Triebfeder* and *Bewegungsgrund* to explain *Motiv*. Whereas in philosophy the former meaning is dominant, the latter is well known to any reader of detective fiction. In his published writings, Kant uses the German *Motiv* only on one occasion: taking up the terminology of his critic Christian Garve in Part I of the essay on *Theory and Practice* (e.g. VIII 282.11).

10 Moreover, the money could become a *Bewegungsgrund* only because of a preceding *desire*; it cannot set human beings in motion just by itself.

valuable action. It is never directed at the effect of an action – which is the proper object of inclination – but always at the action itself (IV 400.19–21). Reverence is linked to the dignity of autonomous legislation in a moral commonwealth at IV 435.17–24 and IV 439.4–12.

As a translation of *Achtung*, 'reverence' seems preferable to 'respect', if only because of the unwelcome connotations the latter word has recently acquired. Today, an expression of 'respect' tends to refer to a – somewhat grudging – concession to another person. We are incessantly urged to feel respect for what divides us, and even for material objects. For Kant, by contrast, *Achtung* is directed at that which all mature and sane human beings share as equal members of the moral community: the faculty of pure practical reason and its law. It is reverence for the law that resides in ourselves and in others that motivates action from duty. Kant himself provides the Latin gloss *reverentia* in the *Metaphysics of Morals* at VI 402.29.

In the *Groundwork*, the notion of reverence for the law is officially introduced in a footnote at IV 401. Kant's argument is, however, rudimentary. There are two reasons for this. First, the *Groundwork*'s declared task of identifying the supreme principle of morality by means of analysing the concept of duty does not require an extensive discussion of human moral psychology. Accordingly, reverence is mentioned only in passing in Sections I and II; and the defence of morality in Section III concerns the metaphysical question of the possibility of moral action rather than its precise mechanism. Secondly, Kant's lectures on moral philosophy bear witness to the fact that for a long time his moral psychology was unstable. If so, it is hardly surprising that the account of reverence in the *Groundwork*, though consistent with the later version in the second *Critique*, should not be fully developed (cf. V 71–89 for a discussion of the intricate psychological mechanism of moral motivation). Note that when at V 76.4–6 Kant says that reverence is not 'the incentive to morality [*Triebfeder zur Sittlichkeit*], but morality itself, subjectively considered as an incentive' he is not denying that reverence motivates moral action. Rather, he is denying that reverence motivates taking up the moral point of view – as every rational agent must – in the first place. Reverence follows once this is done and is then available as an incentive. Kant frequently emphasises the inescapability of reverence, which is a tribute we pay to human decency, whether we like it or not (see e.g. V 76.36–77.5).

VOLITION AND CHOICE

Kant's theory of the will is another terminological minefield. For a start, his earlier writings on moral philosophy require a distinction explicitly made only in the 1790s: the distinction between the will in its lawgiving function (called, like the faculty as a whole, *Wille* in German, *voluntas* in Latin) and the will

in its elective role as the faculty of choice (*Willkür*, *arbitrium*).[11] To complicate matters, the human faculty of volition as a whole is also called the will. The will in its legislative capacity issues the laws that the will in its capacity as the faculty of choice ought to obey. In the *Groundwork*, there is the related notion of a 'pure will' (*reiner Wille*), a will independent of the conditions of human moral psychology whose laws are the subject of a metaphysics of morals (IV 390.35).[12] If the legislative will, to issue a practical decree, requires a preceding inclination, which is rooted in an agent's sensuous nature, the will – as a whole – is heteronomous; it possesses autonomy, by contrast, if it can issue commands of its own (see AUTONOMY AND HETERONOMY, and also FREEDOM).

For a decree of the will to be a command properly speaking the faculty of choice must be able to obey it – otherwise it would not be a command, but a mere 'wish' (see *Anthropology*, VII 251). The will must be able to be practical, i.e. to realise its commands in action. As choice always depends on certain 'incentives' (*Triebfedern*, see MOTIVATION), whereas if free the will in its legislative function may not, there must be a specific incentive directed at doing what is morally required. This interest in doing what the moral law commands is identified with reverence for the law (see REVERENCE). The faculty of the will in its executive function is also identified with the capacity to choose maxims (*Metaphysics of Morals*, VI 226.4–5). The reason is that it is maxims, in conjunction with more specific rules, that determine what action is done in a given situation (see MAXIMS). As it is by virtue of our will that we effect changes in the world, a will is a kind of 'causality' (IV 446.16).

11 The distinction between *arbitrium* and *voluntas* was common in the Latin terminology of his day. Kant occasionally adds the words in brackets to indicate to which function of the will he is referring. Moreover, the notion of a pure will – widely employed by Kant throughout the 1780s – is closely related to *Wille* in its later, narrow sense. It is therefore unlikely that Kant was confused about the will's functions in his earlier writings, even if he failed explicitly to distinguish the two. See e.g. *Critique of Pure Reason* A 534/B 562.

12 For the distinction between a 'holy' and a 'pure' will see the second *Critique*, V 32.17–21.

Bibliography

MAJOR EDITIONS

Kant, Immanuel. *Grundlegung zur Metaphysik der Sitten* (Hartknoch, 1785, [2]1786, reprographic reprint Harald Fischer, 1984)

Grundlegung zur Metaphysik der Sitten, herausgegeben von der Königlich Preußischen Akademie der Wissenschaften, bearbeitet von Paul Menzer in *Kant's gesammelte Schriften*, vol. IV (Reimer, 1903)

Grundlegung zur Metaphysik der Sitten, ed. Karl Vorländer (Dürrsche Buchhandlung, 1906)

Grundlegung zur Metaphysik der Sitten, ed. Artur Buchenau and Ernst Cassirer, in *Kants Werke*, vol. IV (Cassirer, 1921)

Grundlegung zur Metaphysik der Sitten, ed. Rudolf Otto (Klotz, 1930)

Grundlegung zur Metaphysik der Sitten, ed. Theodor Valentiner (Reclam, 1961)

Grundlegung zur Metaphysik der Sitten, ed. Bernd Kraft and Dieter Schönecker (Meiner, 1999)

Grundlegung zur Metaphysik der Sitten, ed. Jens Timmermann (Vandenhoeck & Ruprecht, 2004)

ENGLISH TRANSLATIONS

Kant, Immanuel. *Fundamental Principles of the Metaphysic of Ethics*, trans. Thomas K. Abbott (Longman, 1873)

The Moral Law. Groundwork of the Metaphysic of Morals, trans. H. J. Paton (Hutchinson, 1948)

Foundations of the Metaphysics of Morals, and What is Enlightenment? trans. Lewis White Beck (Bobbs Merrill, 1959)

Grounding for the Metaphysics of Morals, trans. James W. Ellington (Hackett, 1981)

Groundwork of the Metaphysics of Morals, trans. Mary Gregor (Cambridge University Press, 1998; first published in 1996)

Groundwork for the Metaphysics of Morals, trans. Thomas E. Hill Jr and Arnulf Zweig (Oxford University Press, 2002)

Groundwork for the Metaphysics of Morals, trans. Allen W. Wood (Yale University Press, 2002)

Groundwork for the Metaphysics of Morals, trans. Thomas K. Abbott, revised by Lara Denis (Broadview, 2005)

COMMENTARIES AND COLLECTIONS

Aune, Bruce. *Kant's Theory of Morals* (Princeton University Press, 1979)

Duncan, A. R. C. *Practical Reason and Morality. A Study of Immanuel Kant's 'Foundations for the Metaphysics of Morals'* (Thomas Nelson, 1957)

Freudiger, Jürg. *Kants Begründung der praktischen Philosophie* (Paul Haupt, 1993)

Guyer, Paul, ed. *Kant's Groundwork of the Metaphysics of Morals. Critical Essays* (Rowman & Littlefield, 1998)

Höffe, Otfried, ed. *Grundlegung zur Metaphysik der Sitten. Ein kooperativer Kommentar* (Klostermann, 1989, ^{2}1993)

Horn, Christoph and Schönecker, Dieter, eds. *Groundwork for the Metaphysics of Morals* (De Gruyter, 2006)

Paton, H. J. *The Categorical Imperative. A Study in Kant's Philosophy* (Hutchinson, 1947)

Ross, David. *Kant's Ethical Theory. A Commentary on the* Grundlegung zur Metaphysik der Sitten (Clarendon Press, 1954)

Schönecker, Dieter and Wood, Allen W. *Kants 'Grundlegung zur Metaphysik der Sitten'* (Schöningh, 2002)

Schönecker, Dieter. *Kant: Grundlegung III. Die Deduktion des kategorischen Imperativs* (Alber, 1999)

Tittel, Gottlob August. *Über Herrn Kant's Moralreform* (Pfähler, 1786)

Wolff, Robert Paul. *The Autonomy of Reason. A Commentary on Kant's Groundwork of the Metaphysics of Morals* (Harper & Row, 1973)

STUDIES OF KANT'S MORAL PHILOSOPHY

Allison, Henry E. *Kant's Theory of Freedom* (Cambridge University Press, 1990)

Ameriks, Karl. *Interpreting Kant's Critiques* (Clarendon Press, 2003)

Baron, Marcia. *Kantian Ethics almost without Apology* (Cornell University Press, 1995)

Beck, Lewis White. *A Commentary on Kant's Critique of Practical Reason* (University of Chicago Press, 1960)

Cummiskey, David. *Kantian Consequentialism* (Oxford University Press, 1996)

Esser, Andrea. *Eine Ethik für Endliche* (Frommann-Holzboog, 2004)

Gregor, Mary. *Laws of Freedom* (Blackwell, 1963)

Guyer, Paul. *Kant on Freedom, Law, and Happiness* (Cambridge University Press, 2000)

Herman, Barbara. *The Practice of Moral Judgment* (Harvard University Press, 1993)

Hill, Thomas E. *Dignity and Practical Reason in Kant's Moral Theory* (Cornell University Press, 1992)

Human Welfare and Moral Worth. Kantian Perspectives (Clarendon Press, 2002)

Kerstein, Samuel. *Kant's Search for the Supreme Principle of Morality* (Cambridge University Press, 2002)

Korsgaard, Christine. *Creating the Kingdom of Ends* (Cambridge University Press, 1996)

Landucci, Sergio. *Sull'etica di Kant* (Guerini, 1994)

Louden, Robert B. *Morality and Moral Theory* (Oxford University Press, 1992)
Kant's Impure Ethics (Oxford University Press, 2000)
Nell, Onora. *Acting on Principle* (Columbia University Press, 1975)
O'Neill, Onora. *Constructions of Reason* (Cambridge University Press, 1989)
Reath, Andrews. *Agency and Autonomy in Kant's Moral Theory* (Clarendon Press, 2006)
Schwaiger, Clemens. *Kategorische und andere Imperative. Zur Entwicklung von Kants praktischer Philosophie bis 1785* (Frommann-Holzboog, 1999)
Schwartz, Maria. *Der Begriff der Maxime bei Kant* (Lit Verlag, 2006)
Timmermann, Jens. *Sittengesetz und Freiheit. Untersuchungen über Immanuel Kants Theorie des freien Willens* (De Gruyter, 2003)
Timmons, Mark, ed. *Kant's Metaphysics of Morals* (Oxford University Press, 2002)
Tomasi, Gabriele. *Identità razionale e moralità* (Associazione Trentina, 1991)
Wood, Allen W. *Kant's Ethical Thought* (Cambridge University Press, 1999)

OTHER SECONDARY LITERATURE

Allison, Henry E. *Kant's Transcendental Idealism* (Yale University Press, 1983, [2]2004)
Arendt, Hannah. *Eichmann in Jerusalem* (Piper, 1986; first published in 1964)
Barnes, Jonathan. 'Aristotle and the Methods of Ethics', in *Revue Internationale de Philosophie* 34 (1980), 490–511
Bittner, Rüdiger. 'Maximen', in *Akten des 4. Internationalen Kant-Kongresses*, ed. G. Funke (De Gruyter, 1974), vol. II.2, pp. 485–98.
Bittner, Rüdiger and Cramer, Konrad, eds. *Materialien zu Kants 'Kritik der praktischen Vernunft'* (Suhrkamp, 1975)
Cicero. *Abhandlung über die menschlichen Pflichten*, trans. Christian Garve (Wilhelm Gottlieb Korn, 1783)
Cramer, Konrad. *Nicht-reine synthetische Urteile a priori. Ein Problem der Transzendentalphilosophie Immanuel Kants* (Carl Winter, 1985)
Garve, Christian. *Philosophische Anmerkungen und Abhandlungen zu Ciceros Büchern von den Pflichten* (Wilhelm Gottlieb Korn, 1783)
Gilbert, Carlos Melchios. *Der Einfluß von Christian Garves Übersetzung Ciceros 'De Officiis' auf Kants 'Grundlegung zur Metaphysik der Sitten'* (S. Röderer, 1994)
Hare, R. M. 'Could Kant Have Been a Utilitarian', *Utilitas* 5 (1993), 1–16
Henrich, Dieter. 'Das Prinzip der kantischen Ethik', in *Philosophische Rundschau* 2 (1954–5), 29–34
Herman, B. 'Moral Deliberation and the Derivation of Duties', in *The Practice of Moral Judgment* (Harvard University Press, 1993), pp. 132–58
Kagan, Shelly. 'Kantianism for Consequentialists', in Immanuel Kant, *Groundwork for the Metaphysics of Morals*, trans. Allen Wood (Yale University Press, 2002), pp. 111–56
Kuehn, Manfred. *Kant. A Biography* (Cambridge University Press, 2001)
'Kant and Cicero', in *Kant und die Berliner Aufklärung*, ed. V. Gerhardt, R.-P. Horstmann and R. Schumacher (De Gruyter, 2001), vol. III, 270–8

Long, A. A. and Sedley, D. N. *The Hellenistic Philosophers* (Cambridge University Press, 1987)

O'Neill, Onora. 'Kant after Virtue', in *Constructions of Reason* (Cambridge University Press, 1989), pp. 145–62

Patzig, Günther. 'Der Gedanke eines Kategorischen Imperativs', *Archiv für Philosophie* 6 (1956), 82–96

Prauss, Gerold. *Kant und das Problem der Dinge an sich* (Bouvier, 1974, ^{2}1977)

Proops, Ian. 'Kant's Legal Metaphor and the Nature of Deduction', *Journal of the History of Philosophy* 41 (2003), 209–29

Reich, Klaus. *Kant und die Ethik der Griechen* (Mohr, 1935); translated as 'Kant and Greek Ethics' in *Mind* 48 (1939), 446–63

Schiller, Friedrich. *Werke*, ed. G. Fricke and H. G. Göpfert (Hanser, 1987)

Schneewind, J. B. *The Invention of Autonomy* (Cambridge University Press, 1998)

Sidgwick, Henry. *The Methods of Ethics* (Macmillan, 1874, ^{7}1907)

Timmermann, Jens. 'Depositum', *Zeitschrift für philosophische Forschung* 57 (2003), 589–600

'When the Tail Wags the Dog. Animal Welfare and Indirect Duty in Kantian Ethics', *The Kantian Review* 10 (2005), 128–49

'Why Kant Could not Have Been a Utilitarian', *Utilitas* 17 (2005), 243–64

'Kant on Conscience, "Indirect" Duty and Moral Error', *International Philosophical Quarterly* 46 (2006), 293–308

Vaihinger, Hans. *Commentar zu Kants Kritik der reinen Vernunft* (W. Spemann, 1881)

Watkins, Eric. *Kant and the Metaphysics of Causality* (Cambridge University Press, 2005)

Williams, Bernard. 'Internal and External Reasons', in *Moral Luck* (Cambridge University Press, 1981), pp. 101–13 (first published in 1980)

Wolf, Susan. *Freedom within Reason* (Oxford University Press, 1990)

Wood, Allen W. 'The Final Form of Kant's Practical Philosophy', in M. Timmons, ed., *Kant's Metaphysics of Morals* (Oxford University Press, 2002), pp. 1–21

Index

INDEX NOMINUM

INDEX RERUM