Introduction

Jens Timmermann

Its 77 Academy pages make the *Groundwork of the Metaphysics of Morals* a short book, certainly by Kantian standards. Readers of the *Critique of Pure Reason* or the *Metaphysics of Morals* can easily get the impression – true or not – that they are looking at a 'patchwork' of previously existing material. The *Groundwork* is different. It was composed with great care. Moreover, Kant's technical language is absent from the first section, whereas the second employs resounding concepts like that of human beings as 'ends in themselves' or that of a 'kingdom of ends' that we are morally bound to create through our actions. These qualities explain its enduring popularity.

At the same time, the *Groundwork* claims to be as revolutionary in the field of ethics as the *Critique of Pure Reason* was in theoretical philosophy. Kant argues that all other ethical theories are fundamentally unsound because they fail to separate the rational and the natural elements of human volition. An unconditional moral command – a 'categorical imperative' – can only be grounded in pure reason. But this revolution concerns the level of ethical theory, not that of morality. Kant claims to re-establish what he claims are the insights of an uncorrupted common understanding of value and duty against the dangerous perversions peddled by his philosophical opponents. He emphasizes the capacity for self-determination or 'autonomy' that is located within individual human beings; and yet, the law that we impose upon ourselves is not arbitrary, it commands with unrelenting necessity.

As this brief overview indicates, it is hardly an accident that the *Groundwork* has inspired controversy ever since it was first published in 1785. The eleven contributions to this volume show that it still deserves and

The editor would like to thank the Netherlands Institute for Advanced Study for providing outstanding work conditions and support in 2007–8, Lucy Richmond for her help with the manuscript of this volume and an anonymous referee for many helpful comments.

receives careful philosophical, exegetical and historical attention from scholars that belong to remarkably different philosophical traditions.

One of the most striking features of the *Groundwork* is the idea that the final ethical system – a metaphysics of morals – should apply not only to human beings but to all rational beings as such. It is of the utmost necessity, Kant argues in the Preface, to 'work out for once a pure moral philosophy, completely cleansed of everything that may be only empirical and that belongs to anthropology' (*G* IV 389). In the first chapter of this collection ('Ethics and anthropology in the development of Kant's moral philosophy'), Manfred Kuehn traces the tumultuous relation of moral philosophy and the study of human beings through Kant's lecture notes from the 1770s and early 1780s and thus illustrates issues in the development of Kantian ethics in the period between the publication of the *Inaugural Dissertation* in 1770 and the *Groundwork* in 1785. At the time, the anthropology lectures served, at least in part, as an introduction to the moral sciences in general. The emergence of Kant's critical view, Kuehn argues, is of central importance if we want to answer modern critics of Kant's moral philosophy, which see it as too far removed from the complexities of human life. Kant is shown to possess an intricate notion of character; and it turns out that, on closer inspection, the categorical imperative concerns not so much the assessment of token acts but rather the evaluation of practical principles that individual actions merely exemplify. As a result, the pure principles of morality sit comfortably with our everyday manifestations of morality. They do not clash with the reflective views of moral agents – which in the *Groundwork* Kant declares to be his starting point. Kuehn pays particular attention to the impact of Johann Fürchtegott Gellert's thought on the development of Kant's ethical theory.

The next three chapters continue the theme of the moral versus the non-moral, which Kant seeks to define and distinguish on his way to the supreme principle of morality. In her 'Happiness in the *Groundwork*', Alison Hills discusses the moral status of the main rival, ancient and modern, of pure practical reason as the foundation of morality: happiness. In the opening paragraph of the *Groundwork* Kant famously seeks to put happiness in its place. Unlike a morally good will, even happiness, the sum of everything we desire, is not good unconditionally. Happiness is good only if the agent is worthy of it (*G* IV 393). Indeed, for the most part Kant uses happiness and related concepts like inclination and prudential imperatives as a foil to demonstrate what morality is not. At the same time, he concedes that all human beings by nature want to be happy. But what

exactly is happiness? What is the status of prudential reasoning in Kantian moral theory? And how does his ethics of autonomy affect Kant's notion of happiness? Drawing on arguments in Mill and Nagel, Hills suggests that Kant's conception of happiness is an interesting and unusual variation on the standard modern desire-satisfaction model of wellbeing that merits careful philosophical scrutiny.

Perhaps the most notorious thesis put forth in the *Groundwork* is Kant's doctrine that an action possesses moral worth only if it is done solely from duty (e.g. *G* IV 399). In other words: actions motivated by inclination that coincide with moral commands but are not done for the sake of the moral law may be useful, welcome or amiable, but they can never be morally good. In 'Acting from duty: inclination, reason and moral worth', Jens Timmermann examines the underlying assumptions of this view. He argues that 'motivational rigorism' is a consequence of Kant's belief that moral and non-moral volition are different in kind: the latter is directed at some object or state of affairs, whereas the former is directed primarily at volition itself. This radical heterogeneity explains why inclination need not be frustrated if one acts on purely rational grounds (but not vice versa). Consequently, there is no need to relegate the motive of respect to the role of 'backup motive': it makes good sense to say that all moral action must be done for the sake of the law. Also, as morality shapes the life of the virtuous agent without determining the details, there is little danger that Kantian morality will unduly dominate his life. The motivational theory implicit in Section I of the *Groundwork* may thus be more attractive than initially assumed.

At the outset of Section II, Kant turns to examples in moral philosophy, which are portrayed as the central feature of the misguided, haphazard attempts of the popular moral theories of his day. One could not, he says, choose a worse method of moral enquiry than 'wanting to derive it from examples' (*G* IV 408). Examples of virtuous conduct merely illustrate moral principles. It follows that principles are prior. But this just means that examples should be banished from the foundations of ethics, not from moral philosophy as a whole. This is the theme that Robert Louden develops in 'Making the law visible: the role of examples in Kant's ethics'. He examines Kant's objections to examples in the *Groundwork* as well as the place of examples in the moral life of human beings, the distinction between imitation and emulation, and how the teaching of ethics by way of examples – though not as such sufficient – can supplement the cognition and motivational force of an abstract moral law in finite rational beings like ourselves. Drawing on texts such as the *Religion*, the *Metaphysics of Morals*,

the *Anthropology* and Kant's lectures on moral philosophy and education, Louden demonstrates that Kant was not oblivious to the power of example after all.

The three chapters that follow also revolve around the same theme: the formulation of Kant's supreme principle of morals, the categorical imperative, and its various formulations. In 'The moral law as causal law', Robert Johnson examines the connections between rational agency and conformity with universal law, which for Kant is the hallmark of moral volition. After all, the first and basic formulation of the categorical imperative tells us to act only on maxims that we can will as universal laws (*G* IV 421). Johnson argues that the Kantian requirement to conform to laws valid for all rational agents is grounded in the fact not that all agency is *rational* but rather that it is *agency*, i.e. the exercise of a causal power. Whereas reason does not as such provide a universalization requirement, it does provide a spontaneity requirement. According to Johnson, the derivation of the first formulation of the categorical imperative relies on the thought that rational willing is a kind of causation. If so, the claim that the concept of causation contains the idea of conformity to universal laws is not – as many have argued – a trivial claim.

An alternative way of stating the principles of Kantian ethics focuses not on the formal requirement of universalization but on the idea that one should always treat human beings respectfully. This is articulated in the second variant formulation of the categorical imperative to use humanity, whether in one's own person or that of any other, 'always at the same time as an end, never merely as a means' (*G* IV 429). However, the precise reason why one should treat humanity in this manner is highly controversial among Kant scholars. Should we respect human beings because of some absolute inner worth or value that we all possess by virtue of some special capacity or characteristic, our 'dignity', as the standard interpretation has it? Does this turn Kantian ethics into a kind of teleology? Can the requirement to treat others with respect perhaps be justified with recourse to any value at all? In his 'Dignity and the formula of humanity', Oliver Sensen argues for the latter, less common view. One should respect others simply because it is commanded by the categorical imperative. Kant does not ground morality on any value. If so, Kantian ethics can perhaps be shown to be a kind of deontology after all.

Katrin Flikschuh's 'Kant's kingdom of ends: metaphysical, not political' takes issue with a widespread contemporary interpretation of Kant's third variation of the categorical imperative, which involves the notion of an ideal moral commonwealth. On the view in question, the 'kingdom of ends' is a

semi-political entity that represents the normative ideal of a democratic order of mutually legislating and essentially equal citizens; and Kantian moral autonomy is aligned with liberal personal autonomy, understood as the competence of individuals to judge for themselves. In the course of this, the metaphysical elements of Section II of the *Groundwork* are played down or curtailed in the interest of a normative philosophical project. But Kant's 'kingdom of ends', Flikschuh argues, is an essentially metaphysical conception of a non-political order with God at the top that cannot be appropriated by political theory without distorting both Kantian ethics and Kant's basic doctrine of right, which concerns the much more confined notion of *Rechtsstaat*.

Jerome Schneewind's contribution returns to the historical context of Kant's *Groundwork of the Metaphysics of Morals*. Towards the end of Section II, as the categorical imperative has duly been revealed as a principle of autonomy, Kant rather unflatteringly lumps together the views of earlier moral philosophers and dismisses them because they are, one and all, expressions of the misguided assumption of heteronomy (*G* IV 441). In 'Kant against the "spurious principles of morality"' Schneewind considers the claim that Kant's argument for the categorical imperative as the supreme principle of morality succeeds in dismissing all rival principles by means of an 'only survivor' argument without begging the question. The final verdict is negative. Whereas Kant's objections to egoistic hedonism, Wolffian perfectionism and Crusius' divine command theory are telling, in other cases he has to rely on assumptions that his opponents are supposed to share while in fact they do not; there are some contemporary views that he does not consider, and more recent alternatives were not, of course, available to him to be dismissed in this fashion. Nonetheless, Schneewind concludes that only Hume matches Kant in trying to give fair-minded and often trenchant responses to the main rivals of his own theory.

The final three contributions again concern a single theme: that of the ambitious Kantian project contained in the last section of the book. John Skorupski discusses Kant's 'grand claim' that we can derive from the notion of free or rational action a principle of how all rational beings should act, which he equates with a peculiar kind of impartiality: the categorical imperative ('Autonomy and impartiality'). He breaks the Kantian argument down into two steps: the first leads from the notion of acting on a reason to autonomy, the second from autonomy to the impartiality of Kantian morality. If successful, this move would help us against the moral sceptic, who is unlikely to doubt the existence of practical reason. But whereas the

first step is defensible, Skorupski argues, the second one fails: morality cannot be derived analytically from the mere idea of autonomy. If so, the 'grand project' pursued in the third section of the *Groundwork* fails.

Paul Guyer's 'Problems with freedom: Kant's Argument in *Groundwork* III and its subsequent emendations' centres on Kant's claim that Section III provides a rudimentary 'critique of the subject, that is, of pure practical reason' (*G* IV 440), which is meant to substantiate the analytic claims of Sections I and II. This critique, Guyer claims, is supposed to achieve this end by means of a metaphysical argument that depends on a claim about our real, 'noumenal' selves, and that is intended to prove that the moral law is a causal law of the 'real' self. He continues by showing how the problems raised by this conception of freedom occupy Kant in his later writings, notably the *Critique of Practical Reason* and *Religion Within the Limits of Mere Reason*.

By contrast, Fred Rauscher argues against the prevalent reading of the deduction of the categorical imperative in Section III, which assumes that Kant is trying to provide an argument for the validity of the moral law for human beings by drawing on a theoretical argument about the nature of reality – transcendental idealism – borrowed from the *Critique of Pure Reason* ('Freedom and reason in *Groundwork* III'). He contrasts this reconstruction with his own 'validation of reason' interpretation, which does not pretend to provide proof of the objective reality of morality but rather seeks to explain only the inevitability of the ascription of morality to human beings who take themselves to be rational agents. He argues that Kant invokes transcendental freedom not as a feature of the whole person or the choice, but only of the faculty of reason that we all possess as a way of explaining freedom of the will. In doing so, Rauscher emphasizes the limitations of the project of deduction that Kant himself stresses on the last page of the *Groundwork* (*G* IV 463).

CHAPTER 1

Ethics and anthropology in the development of Kant's moral philosophy

Manfred Kuehn

In the *Groundwork of the Metaphysics of Morals* of 1785, Immanuel Kant sharply distinguishes moral philosophy from anthropology, claiming that the metaphysics of morals must precede practical anthropology, must be completely *a priori*, and must therefore be 'purified' or 'cleansed' of anything empirical, *a posteriori*, or belonging to mere anthropology (*G* IV 388).[1] The rudiments of the *a priori* moral philosophy expounded in the *Groundwork* are extensively explained and analysed in the two other major works that are explicitly concerned with moral philosophy, the *Critique of Practical Reason* (1787) and the *Metaphysics of Morals* (1797). The *a posteriori* or empirical doctrine of morals or what is 'called more specifically practical anthropology', by contrast, never really comes into focus in the published works. Even the *Anthropology from a Pragmatic Point of View* contains little that would directly concern moral philosophy. Reflections on the empirical and anthropological aspects of morality can, however, be found in the student notes of his lectures on moral philosophy given before 1785, that is, before the publication of the *Groundwork*, and, more importantly, in those of the lectures on anthropology, which Kant began to offer in the winter semester of 1772–73 with the hope of transforming the subject into a proper academic discipline. Kant's main goal in this new *collegium privatim* was, as he put it, to 'introduce all the sciences that are concerned with morals, with the ability of commerce, and the method of educating and ruling human beings, or all that is practical' (X 145). As such a general introduction into practical philosophy, the anthropology lectures had the closest relation to moral philosophy, which Kant always offered during the winter semesters. The anthropology lectures were conceived, at least in part, as a general introduction to or preparation for moral philosophy. But even if

[1] Compare also *MdS* VI 247: 'If there is any subject matter that allows of *philosophy* (a system of rational cognitions on the basis of concepts), then there must exist for this philosophy also a system of pure rational concepts independent of the conditions of intuition, i.e. a metaphysics.'

it were argued that Kant himself did not see the relation of anthropology and moral philosophy in this way, it is clear that there must have been some overlap and cross-fertilization between these two subjects taught so closely together.[2]

These early lectures contain, therefore, just the notions that characterized for Kant the kind of morality that was not yet 'carefully cleaned of everything empirical'. Therefore, a closer analysis of these lectures seems to me to be of central importance for understanding pure moral concepts, for we may assume that even something that has carefully been cleaned still has the contours of the object not yet cleaned or still encrusted with impurities. At the very least, one should expect that the cleaned object does not contain entirely new or different features than the one that has not as yet been cleaned. Furthermore, the cleaning process should only remove the impurities, not parts of the object to be cleaned; nor, one might argue, can the cleaned object have a shape that could not find its place within the object encrusted with impurities. In other words, the pure principles of morality must fit with the empirical manifestations of morality. Pure moral philosophy should not have a 'shape' entirely different from what most people would consider moral. If we take Kant's metaphor seriously, there should be no incompatibility between his metaphysics of morals and morality.

Many philosophers have argued just that. Kant's metaphysics of morals seriously misconstrues the moral domain. Bernard Williams is perhaps the most important of those who have done so recently.[3] Others have argued that Kant's moral philosophy does not only not involve fundamental metaphysical assumptions, but captures very well our common preconception of morality.[4] However, before we decide whether or not Kant's modern critics or defenders are right, it might be of interest to see what 'empirical concepts' of morality Kant actually started out from, and what it was that he thought it was necessary to cleanse or purify. If only for this reason, it should be rewarding to investigate in some detail the kind of morality from which Kant started out. In any case, this is what I intend to do in this chapter. I would like to investigate what the contents of Kant's anthropology lectures between 1772 and 1785 show about the origins of some of the

[2] Kant himself refers his students in the ethics lectures to the anthropology lectures. See XXVII 466, for instance.

[3] See, e.g., Bernard Williams, *Ethics and the Limits of Philosophy* (Cambridge University Press, 1985), pp. 174, 180, 190f.

[4] See, e.g., Christine M. Korsgaard, *The Sources of Normativity* (Cambridge University Press, 1996) or Barbara Herman, *The Practice of Moral Judgment* (Harvard University Press, 1993), or any number of contemporary American 'Kantians'.

central concepts of the *Groundwork*, hoping to illuminate certain issues of the development of his moral philosophy between, that is, the period that roughly lasted from the publication of the *Inaugural Dissertation* (1770) to the *Groundwork of the Metaphysics of Morals* (1785).[5] These issues concern the closely related concepts of 'moral sense', 'moral character', 'maxim' and 'the good will'.

Especially 'character' has received some attention lately. Barbara Herman, for instance, has pointed out in an influential paper 'some of the resources that might be drawn on to develop a Kantian idea of character and to indicate … some of its advantages for moral judgment'.[6] Others, like Otfried Höffe, have argued for a relevance of Aristotle in this context.[7] Still others, like Nancy Sherman, have tried to show, while being 'more faithful to the texts' and 'responsive to debates in contemporary ethics', that Kant and Aristotle are closer on character because emotions actually play a central role for Kant as well.[8] While I will not directly engage these proposals here, it should be clear to anyone acquainted with the recent literature that my discussion provides an alternative to such views.

[5] Though I have written on Kant's moral development several times before, this chapter is very different in focusing more explicitly on the anthropology. See 'The Moral Dimension of Kant's Inaugural Dissertation: A New Perspective on the "Great Light of 1769?"' in *Proceedings of the 8th International Kant Congress in Memphis* (Milwaukee: Marquette University Press, 1995), vol. 1.2, pp. 373–92; 'Kant and Cicero' in Volker Gerhardt, Rolf-Peter Horstmann and Ralph Schumacher (eds.), *Proceedings of the 9th International Kant Congress in Berlin, April 2000* (Berlin/New York: de Gruyter, 2001), pp. 270–8; 'Einleitung' in Immanuel Kant, *Vorlesungen zur Moralphilosophie* (Werner Stark (ed.), Berlin: de Gruyter, 2004), pp. vii–xxxv; 'Introduction' in Immanuel Kant, *Anthropology from a Pragmatic Point of View* (Robert Louden (ed.), Cambridge University Press, 2006), pp. vii–xxxiii.

[6] Barbara Herman, 'Making Room for Character' in Stephen Engstrom and Jennifer Whiting (eds.), *Aristotle, Kant and the Stoics: Rethinking Happiness and Duty* (Cambridge University Press, 1996), pp. 36–62 at p. 37. Herman is responding in this context to McDowell, who claims that virtue is also a perceptive ability that allows us to see 'situations in a certain distinctive way' and make moral judgements. See John McDowell, 'Virtue and Reason', *The Monist* 62 (1979), 331–50.

[7] Otfried Höffe, 'Universalistische Ethik und Urteilskraft: ein aristotelischer Blick auf Kant', *Zeitschrift für philosophische Forschung* (1990), 537–63 and his 'Aristoteles' universalistische Tugendethik' in Klaus Peter Rippe and Peter Schaber (eds.), *Tugendethik* (Stuttgart: Reclam, 1998), pp. 42–68, which seems to be influenced very much by Martha Nussbaum's 'Non-Relative Virtues: an Aristotelian Approach' in Peter A. French, Theodore E. Uehling, Jr and Howard K. Wettstein (eds.), *Ethical Theory: Character and Virtue*, Midwest Studies in Philosophy 13 (Notre Dame, IN: University of Notre Dame Press, 1988), pp. 32–53, an expanded version of which is in 'Non-Relative Virtues: an Aristotelian Approach' in Martha C. Nussbaum and Amartya Sen (eds.), *The Quality of Life, Papers presented at a conference sponsored by the World Institute for Development Economics Research*, WIDER Studies in Development Economics (Oxford: Clarendon Press; New York: Oxford University Press, 1993), pp. 242–69.

[8] Nancy Sherman, *Making a Necessity of Virtue: Aristotle and Kant on Virtue* (Cambridge University Press, 1997), especially pp. 121–86.

1 FROM MORAL SENSE TO MORAL CHARACTER: CHRISTIAN FÜRCHTEGOTT GELLERT AND 'IDLE IDEALS AND DESIRES' VERSUS 'GUTE DENKUNGSART'

It is well known that during the 1760s, Kant was very much influenced by Francis Hutcheson and considered the moral sense or 'moral feeling', as he and others also called it, as the foundation of morality.[9] It is also well known that he changed his mind some time after 1764, that is, after the publication of his Prize Essay for the Berlin Academy: *Inquiry concerning the Distinctness of the Principles of Natural Theology and Morality*. In that work, Kant still found that Hutcheson has provided 'the starting point for some nice observations about what he called the moral feeling' and leaves open the possibility that the 'first principles' of obligation may be found in such a feeling. If we can trust Adickes' dating of the reflections, then Kant had already abandoned by 1770 or 1771 the view that moral feeling or moral sense could play any foundational role.[10] For in the reflections of that period he found that the real question of morality is (XIX 135):

> whether moral judgements are concerned with whether the actions [judged] are seen as good or as pleasant. If it is the former, then it is the quality of the action which is identical for every understanding that forms the basis of the judgement, and this is the effect of reason. If it is the second, then one judges on the basis of feeling, and this is not necessarily valid for everyone.

In these reflections, he also suggested that the moral sense may not be an original sense, but rather something derivative and instinctual. It does not lead to moral judgements, but to inclinations. Furthermore, he thought that the moral sense needs to be formed by education or by 'concepts and rules' (XIX 137).

This view can also be found in the lecture notes on anthropology called *Collins*, which were taken during the winter semester 1772–73. In a section that is concerned with pleasure and displeasure arising from the beautiful

[9] Thus Henrich has claimed that 'Kant became aware of the general situation of ethics at the middle of the eighteenth century through the opposition between Wolff's *philosophia practica universalis* and Hutcheson's moral philosophy, and his first independent formulation of an ethical theory resulted from a critique of these two philosophers' (Dieter Henrich, 'The Concept of Moral Insight' in *The Unity of Reason: Essays on Kant's Philosophy* (R. Velkley (ed.), Manfred Kuehn (trans.), Harvard University Press, 1994), pp. 55–88. Apart from the fact that there was no thorough opposition, but rather a perceived complementarity, this claim situates Kant's early theory correctly. See also Dieter Henrich, 'Hutcheson und Kant', *Kant-Studien* 49 (1957/58), 49–69, and 'Über Kants früheste Ethik', *Kant-Studien* 54 (1963), 404–31.

[10] See XXV 1–228. These notes are based mainly on a set of notes taken by Collins, but they are supplemented by materials from lecture notes, taken by others.