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).